Asymptotic statistical equivalence for ergodic diffusions: the multidimensional case

Arnak S. Dalalyan, Markus Reiss

Asymptotic statistical equivalence for ergodic diffusions: the multidimensional case

Arnak Dalalyan[1] · Markus Reiß[2]

May 6, 2005

Key words. Asymptotic equivalence, statistical experiment, Le Cam distance, ergodic diffusion, Gaussian shift, heteroskedastic regression

Mathematics Subject Classification (2000): 62B15, 62G05, 62G07, 62G20, 62M05.

Abstract. Asymptotic local equivalence in the sense of Le Cam is established for inference on the drift in multidimensional ergodic diffusions and an accompanying sequence of Gaussian shift experiments. The nonparametric local neighbourhoods can be attained for any dimension, provided the regularity of the drift is sufficiently large. In addition, a heteroskedastic Gaussian regression experiment is given, which is also locally asymptotically equivalent and which does not depend on the centre of localisation. For one direction of the equivalence an explicit Markov kernel is constructed.

1. Introduction

Asymptotic equivalence is a powerful concept for analysing statistical inference problems by a transfer to the analogous problem in a simpler statistical experiment. A breakthrough were the results by Brown and Low [5] and Nussbaum [18] who established asymptotic equivalence of the two classical experiments, one-dimensional Gaussian regression and density estimation, with an accompanying sequence of Gaussian shift experiments. In this paper we consider the statistical inference for the drift in a multidimensional diffusion experiment under stationarity assumptions and prove the asymptotic equivalence with corresponding multidimensional Gaussian shift and regression experiments.

Asymptotic equivalence results for dependent data are not very numerous, see Dalalyan and Reiß [10] for an overview. Even for simple experiments, as

Laboratoire de Probabilités, Université Paris VI, Place Jussieu, 75252 Paris Cedex 05, France, e-mail: dalalyan@ccr.jussieu.fr

Weierstraß Institute for Applied Analysis and Stochastics, Mohrenstraße 39, 10117 Berlin, Germany, e-mail: mreiss@wias-berlin.de

the classical ones described above, results for asymptotic equivalence in the multidimensional case are very scarce. We only know of the recent work by Carter [8] who proves asymptotic equivalence for two-dimensional Gaussian regression, but argues that his method fails for higher dimensions. One of the main reasons for the difficulties in transferring methods to higher dimensions is that piecewise constant approximations of the unknown functional parameter usually do not suffice anymore and higher order approximations have to be used, which creates unexpected problems. Brown and Zhang [6] remark that the two classical experiments and their accompanying Gaussian shift experiments are not asymptotically equivalent in the case of nonparametric classes of Hölder regularity $\beta \le d/2$, where d denotes the dimension.

The methodology we applied in [10] to establish asymptotic equivalence for scalar diffusions relied heavily on the concept of local time. For multidimensional diffusions local time does not exist. This might explain why the statistical theory for scalar diffusions is very well developed (see Kutoyants [15]), while inference problems for multidimensional diffusions are more involved and much less studied. We refer to Bandi and Moloche [2] for the analysis of kernel estimators for the drift vector and the diffusion matrix and to Aït-Sahalia [1] for a recent discussion of applications for multidimensional diffusion processes in econometrics.

In Section 2 we review results for multidimensional diffusions and construct estimators for the invariant density and the drift vector. Interestingly, the estimator of the invariant density converges for $d \ge 2$ with a rate which is slower than parametric, but faster than in classical d-dimensional density estimation problems. The local equivalence result of the multidimensional diffusion experiment with an accompanying Gaussian shift experiment is formulated and described in Section 3. The local neighbourhoods can be attained for drift functions in a nonparametric class of regularity $\beta > (d-1+\sqrt{2(d-1)^2-1})/2$ for any dimension $d \ge 2$. In Section 4 the corresponding equivalence with a heteroskedastic regression experiment, which does not depend on the centre of localisation, is treated. This can be used to establish global equivalence with a single experiment, which even in the one-dimensional case cannot be obtained for the Gaussian shift experiment due to the absence of a variance stabilising transform, as was first noted by Delattre and Hoffmann [11]. The explicit construction of a Markov kernel establishing the important part of the asymptotic equivalence is presented in Section 5. The proof of the main local equivalence result is deferred to Section 6.

2. Preliminaries

2.1. Diffusion processes

We assume that a continuous record $X^T = \{X_t,\ 0 \le t \le T\}$ of a d-dimensional diffusion process X is observed up to time instant T. This diffusion process is supposed to be given as a solution of the stochastic differential equation

$$dX_t = b(X_t)\,dt + dW_t, \qquad X_0 = \xi, \qquad t \in [0,T], \tag{1}$$

where $b : \mathbb{R}^d \to \mathbb{R}^d$, $W = (W_t,\, t \ge 0)$ is a d-dimensional Brownian motion and ξ is a random vector independent of W. We denote by $b_i : \mathbb{R}^d \to \mathbb{R}$, $i = 1, \dots, d$, the components of the vector valued function b. In what follows, we assume that the drift is of the form $b = -\nabla V$, where $V \in C^2(\mathbb{R}^d)$ is referred to as potential. This restriction permits to use strong analytical results for the Markov semigroup of the diffusion on the L^2-space generated by the invariant measure.

For positive constants M_1 and M_2, we define $\Sigma(M_1, M_2)$ as the set of all functions $b = -\nabla V : \mathbb{R}^d \to \mathbb{R}^d$ satisfying for any $x, y \in \mathbb{R}^d$

$$|b(x)| \le M_1(1 + |x|), \tag{2}$$

$$(b(x) - b(y))^T(x - y) \le -M_2|x - y|^2, \tag{3}$$

where $|\cdot|$ denotes the Euclidian norm in $\mathbb{R}^d$. Any such function b is locally Lipschitz-continuous. Therefore equation (1) has a unique strong solution, which is a homogeneous continuous Markov process, cf. Rogers and Williams [22], Thm. 12.1. Set $C_b = \int_{\mathbb{R}^d} e^{-2V(u)}\,du$ and

$$\mu_b(x) = C_b^{-1} e^{-2V(x)}, \qquad x \in \mathbb{R}^d.$$

Under condition (3) we have $C_b < \infty$ and the process X is ergodic with unique invariant probability measure (Bhattacharya [3, Thm. 3.5]). Moreover, the invariant probability measure of X is absolutely continuous with respect to the Lebesgue measure and its density is μ_b. From now on, we assume that the initial value ξ in (1) follows the invariant law such that the process X is strictly stationary. We denote by $\mathbf{P}_b^T$ the law of this process induced on the canonical space $\big(C([0,T];\mathbb{R}^d), \mathcal{B}_{C([0,T];\mathbb{R}^d)}\big)$ and by $\mathbf{E}_b$ the expectation operator with respect to this law. We write $\mu_b(f) := \mathbf{E}_b[f(X_0)] = \int f\mu_b$. Let $P_{b,t}$ be the transition semigroup of this process on $L^2(\mu_b)$, that is

$$P_{b,t}f(x) = \mathbf{E}_b[f(X_t)|X_0 = x],\ f \in L^2(\mu_b) = \Big\{f : \mathbb{R}^d \to \mathbb{R} : \int |f|^2\mu_b < \infty\Big\}.$$

The transition density is denoted by $p_{b,t}$: $P_{b,t}f(x) = \int f(y)p_{b,t}(x,y)\,dy$.

2.2. Estimators of drift and invariant density

Some notation. We write $A(p) \lesssim B(p)$ when $A(p)$ is bounded by a constant multiple of $B(p)$ uniformly over the parameter values p, that is $A(p) = \mathcal{O}(B(p))$ using the Landau symbol. Similarly, $A(p) \sim B(p)$ means that $A(p) \lesssim B(p)$ as well as $B(p) \lesssim A(p)$. We denote by $|A|$ the Lebesgue measure and by $\operatorname{diam}(A)$ the diameter of a Borel set $A \subset \mathbb{R}^d$.

For any multi-index $\alpha \in \mathbb{N}^d$ and $x \in \mathbb{R}^d$ we set $|\alpha| = \alpha_1 + \ldots + \alpha_d$ and $x^\alpha = x_1^{\alpha_1} \cdot \ldots \cdot x_d^{\alpha_d}$. Let us introduce the Hölder class

$$\mathcal{H}(\beta, L) = \left\{ f \in C^{\lfloor \beta \rfloor}(\mathbb{R}^d; \mathbb{R}) \,:\, \begin{array}{c} |D^\alpha f(x) - D^\alpha f(y)| \le L|x-y|^{\beta - \lfloor \beta \rfloor} \\ \text{for any } \alpha \text{ such that } |\alpha| = \lfloor \beta \rfloor \end{array} \right\}$$

where $\lfloor \beta \rfloor$ is the largest integer *strictly* smaller than β and $D^\alpha f := \frac{\partial^{|\alpha|} f}{\partial x_1^{\alpha_1} \ldots \partial x_d^{\alpha_d}}$.

The construction. Let us assume that the potential V lies in $\mathcal{H}(\beta+1, L)$ for some β, $L > 0$, which implies $b_i \in \mathcal{H}(\beta, L)$. Furthermore, if for some constant $C_1 > 0$ we have

$$\max_{i=1,\ldots,d} \max_{\alpha:|\alpha| \le \lfloor \beta \rfloor} |D^\alpha b_i(0)| \le C_1 \tag{4}$$

then the function μ_b is Hölder continuous of order $\beta + 1$ in any bounded set $A \subset \mathbb{R}^d$, that is

$$|D^\alpha \mu_b(x) - D^\alpha \mu_b(y)| \le L_\mu |x-y|^{\beta - \lfloor \beta \rfloor}, \qquad \forall \alpha \in \mathbb{N}^d : |\alpha| = \lfloor \beta \rfloor + 1$$

for all $x, y \in A$ and for some constant L_μ. We denote by $\widetilde{\mathcal{H}}(\beta, L, C_1)$ the set of all functions b such that $b_i \in \mathcal{H}(\beta, L)$ and (4) is fulfilled.

A natural kernel estimator for the invariant density based on the observation X^T is given by

$$\hat{\mu}_{h,T}(x) = \frac{1}{T} \int_0^T K_h(x - X_t)\, dt, \quad x \in \mathbb{R}. \tag{5}$$

Here, $K_h(x) = h^{-d} K(h^{-1}x)$ and $K : \mathbb{R}^d \to \mathbb{R}$ is a smooth kernel function of compact support, satisfying $\int K(x)\, dx = 1$ and $\int K(x) x^\alpha\, dx = 0$ whenever $1 \le |\alpha| \le \lfloor \beta \rfloor + 1$. The usual bias-variance decomposition and approximation inequality yield (Efromovich [12], § 8.9)

$$\mathbf{E}_b\big[|\hat{\mu}_{h,T}(x) - \mu_b(x)|^2\big] \lesssim h^{2(\beta+1)} + T^{-2} \operatorname{Var}\Big[\int_0^T K_h(x - X_t)\, dt\Big]. \tag{6}$$

By analogy with the model of regression with random design, a reasonable estimator of b is obtained by setting

$$\hat{b}_{h,T}(x) = \frac{\int_0^T K_h(x - X_t)\, dX_t}{T \max(\hat{\mu}_{h,T}(x), \mu_*(x))}, \quad x \in \mathbb{R}, \tag{7}$$

where $\mu_*(x) > 0$ is some a priori lower bound on $\mu_b(x)$, see Remark 6 below. A similar risk analysis gives for $i = 1, \dots, d$:

$$\begin{aligned}\mathbf{E}_b\big[|\hat{b}_{i,h,T}(x) - b_i(x)|^2\big] \lesssim h^{2\beta} + \frac{1}{Th^d} + \frac{1}{T^2} \operatorname{Var}\Big[\int_0^T K_h(x - X_t) b_i(X_t)\, dt\Big] \\ + \mathbf{E}_b\big[|\hat{\mu}_{h,T}(x) - \mu_b(x)|^2\big].\end{aligned} \tag{8}$$

Asymptotic results. In order to determine the asymptotic behaviour for $T \to \infty$, we study the variance of general additive functionals of X in d dimensions. To do so, we assume that the semigroup $P_{b,t}$ enjoys the following properties.

Assumption 1 (spectral gap inequality) *There exists a $\rho > 0$ such that for any $f \in L^2(\mu_b)$ and for any $t > 0$*

$$\|P_{b,t} f - \mu_b(f)\|_{\mu_b} \leq e^{-t\rho} \|f\|_{\mu_b}.$$

Assumption 2 *There is a $C_0 > 0$ such that for any $t > 0$ and for any pair of points $x, y \in \mathbb{R}^d$, satisfying $|x - y|^2 < t$, we have*

$$p_{b,t}(x, y) \leq C_0(t^{-d/2} + t^{3d/2}).$$

Remark 1. Due to Remark 4.14 in Chen and Wang [9] Assumption 1 is fulfilled with $\rho = M_2$, whenever (3) holds.

Remark 2. If b fulfills (2), then Assumption 2 can be deduced from Qian and Zheng [20, Thm. 3.2]. Indeed, taking in that inequality $q = 1 + t$ and bounding the terms ζ_q and ρ_q respectively by $Cq^{3/2}$ and Cq, we get the desired inequality. If moreover b is bounded, Assumption 2 is satisfied for every $(x, y) \in \mathbb{R}^d$ and without the term $t^{3d/2}$ at the right-hand side, cf. Qian *et al.* [19, inequality (5)].

Proposition 1. *Let r be a positive number and $f : \mathbb{R}^d \to \mathbb{R}$ be a bounded, measurable function with support $\mathcal{S}$ satisfying* $\operatorname{diam}(|\mathcal{S}|)^d < r^d |\mathcal{S}|$ *and* $|\mathcal{S}| < 1$*. Under Assumptions 1 and 2 there exists a constant C depending only on r, $d \geq 2$ and on C_0 and ρ from Assumptions 1 and 2 such that*

$$\operatorname{Var}_b\left(\int_0^T f(X_t)\, dt\right) \leq CT \|f\|_\infty^2 \mu_b(\mathcal{S}) |\mathcal{S}| \psi_d^2(|\mathcal{S}|),$$

where $\|f\|_\infty = \sup_{x\in\mathbb{R}^d} |f(x)|$ *and*

$$\psi_d(x) = \begin{cases} \max(1, (\log(1/x))^2), & d = 2, \\ x^{1/d-1/2}, & d \geq 3. \end{cases}$$

Proof. Set $f_c = f - \mu_b(f)$. Symmetry and stationarity yield

$$\begin{aligned} \mathrm{Var}_b\left(\int_0^T f(X_t)\,dt\right) &= 2\int_0^T\int_0^s \mathbf{E}_b\big[f_c(X_t)f_c(X_s)\big]\,dt\,ds \\ &= 2\int_0^T\int_0^s \mathbf{E}_b\big[f_c(X_0)f_c(X_{s-t})\big]\,dt\,ds \\ &= 2\int_0^T (T-u)\mathbf{E}_b\big[f_c(X_0)f_c(X_u)\big]\,du \\ &\leq 2T\int_0^T \big\langle f_c, P_{b,u}f_c\big\rangle_{\mu_b}\,du. \end{aligned}$$

Let $0 < \delta < D \leq T$ where the specific choice of δ, D is given later. Then

$$\int_{[0,\delta]\cup[D,T]} \big\langle f_c, P_{b,u}f_c\big\rangle_{\mu_b}\,du \leq (\delta + \rho^{-1}e^{-\rho D})\|f\|_{\mu_b}^2 \lesssim (\delta + e^{-\rho D})\mu_b(\mathcal{S})\|f\|_\infty^2 \tag{9}$$

follows from $\|P_{b,u}f_c\|_{\mu_b} \leq e^{-\rho u}\|f\|_{\mu_b}$ given by Assumption 1. For moderate values $u \in [\delta, D]$ we use

$$\langle f_c, P_{b,u}f_c\rangle_{\mu_b} \leq \langle f, P_{b,u}f\rangle_{\mu_b} \leq \int |f(x)|\Big(\int p_{b,u}(x,y)\,|f(y)|\,dy\Big)\,\mu_b(x)\,dx.$$

For $\delta > \mathrm{diam}(\mathcal{S})^2$ we infer from Assumption 2

$$\langle f, P_{b,u}f\rangle_{\mu_b} \leq C(u^{-d/2} + u^{3d/2})\mu_b(|f|)\int |f(y)|\,dy \quad \forall\, u \geq \delta. \tag{10}$$

Combining (9) and (10) and assuming $\mathrm{diam}(\mathcal{S}) < \delta^{1/2}$, for $d > 2$ we find

$$\int_0^T \big\langle f_c, P_{b,u}f_c\big\rangle_{\mu_b}\,du \lesssim \Big(\delta + e^{-\rho D} + \delta^{1-d/2}|\mathcal{S}| + D^{1+3d/2}|\mathcal{S}|\Big)\mu_b(\mathcal{S})\|f\|_\infty^2.$$

Balancing the terms, we choose $D = \max(-\rho^{-1}\log(|\mathcal{S}|), r^2)$ and $\delta = r^2|\mathcal{S}|^{2/d}$. This gives the asserted estimate because we had assumed $\mathrm{diam}(\mathcal{S}) < r|\mathcal{S}|^{1/d}$. The case $d = 2$ can be treated similarly. □

Remark 3. In the case $d = 1$ the bound holds with $\psi_1(x) = 1$, cf. Proposition 5.1 in Dalalyan and Reiß [10].

Remark 4. The dimensional effect is due to the singular behaviour of $p_{b,t}(x,y)$ for $t \to 0$. However, if the term $t^{3d/2}$ is absent in Assumption 2, then in the definition of ψ_2 the term $(\log(1/|\mathcal{S}|))^2$ can be replaced by $(\log(1/|\mathcal{S}|))^{1/2}$. This is the case when the drift is bounded.

Corollary 1. *If $b \in \widetilde{\mathcal{H}}(\beta, L, C_1) \cap \Sigma(M_1, M_2)$, the estimators given in* (5) *and* (7) *satisfy for h sufficiently small the following risk estimates:*

$$\mathbf{E}_b\left[(\hat{\mu}_{h,T}(x) - \mu_b(x))^2\right] \lesssim h^{2(\beta+1)} + T^{-1}\psi_d^2(h^d),$$
$$\mathbf{E}_b\left[|\hat{b}_{h,T}(x) - b(x)|^2\right] \lesssim h^{2\beta} + T^{-1}h^{-d} + h^{2(\beta+1)} + T^{-1}\psi_d^2(h^d).$$

The rate-optimal choice $h = h(T) \sim T^{-1/(2\beta+d)}$ yields the rates

$$\mathbf{E}_b\left[(\hat{\mu}_{h(T),T}(x) - \mu_b(x))^2\right]^{1/2} \lesssim \begin{cases} T^{-1/2}(\log T)^2, & d = 2, \\ T^{-(\beta+1)/(2\beta+d)}, & d \geq 3, \end{cases}$$
$$\mathbf{E}_b\left[|\hat{b}_{h(T),T}(x) - b(x)|^2\right]^{1/2} \lesssim T^{-\beta/(2\beta+d)}.$$

Proof. The risk bound for $\hat{\mu}_{h,T}$ follows from $|\operatorname{supp}(K_h)| \sim h^d$, $\|\mu_b\|_\infty \lesssim 1$ and an application of Proposition 1 to the bias-variance decomposition (6) for any h sufficiently small. In the same way, we obtain the estimate for each $\hat{b}_{i,T,h}$ and the rates follow by simple substitution. $\square$

Remark 5. The convergence rates for the risk of $\hat{\mu}$ are to be compared with the one-dimensional case, where the parametric rate $T^{-1/2}$ is obtained, and with standard multivariate density estimation, where the corresponding rate is $n^{-\beta/(2\beta+d)}$ for n observations, which is considerably larger. In contrast, the rate for $\hat{b}$ corresponds exactly to the classical rate $n^{-\beta/(2\beta+d)}$ in regression or density estimation.

Remark 6. Using conditions (2), (3) and the equality $V(x) = V(0) - \int_0^1 b(tx)^T x\,dt$, we find

$$-M_1|x| + \frac{1}{2}M_2|x|^2 \leq V(x) - V(0) \leq \frac{1}{2}M_1|x|^2 + M_1|x|.$$

Therefore, we can take $\mu_*(x) = e^{-M_1|x|^2 - 2M_1|x|} / \int e^{2M_1|y| - M_2|y|^2} dy$ as an a priori lower bound for $\mu_b(x)$. Moreover, due to assumption (4) the function μ_b is Hölder continuous in $A_\delta = \{x \in \mathbb{R}^d : \inf_{y\in A} |x - y| \leq \delta\}$ for any $\delta > 0$ and for any bounded set $A \subset \mathbb{R}^d$. Therefore we do not need to modify the kernel estimators at the boundaries of A and the inequalities of Corollary 1 hold uniformly in b and in $x \in A$.

Remark 7. Corollary 1 describes the rates of convergence of estimators for the local risk, that is for a pointwise loss function. To attain the local neighbourhood defined in the next section, the risk given by the sup-norm loss must be studied. In the classical problems of nonparametric estimation, the rates of convergence for the sup-norm loss on a compact set coincide up to a logarithmic factor with the local rates of convergence (Korostelev and Nussbaum [14], Giné, Koltchinskii and Zinn [13]). The extension from the pointwise to the uniform loss result is usually fairly standard, but more involved and lies out of the scope of this paper.

3. Equivalence with the Gaussian shift model

3.1. Statement of the result

Let $\Sigma_\beta(L, M_1, M_2)$ be the set of functions $b \in \Sigma(M_1, M_2)$ such that all d components b_i of b are in $\mathcal{H}(\beta, L)$. We fix a function $b^\circ \in \Sigma_\beta(L, M_1, M_2)$. Our main result establishes a local asymptotic equivalence between diffusion and Gaussian shift models in the local setting, that is when the parameter set is a shrinking neighbourhood of b°. $\mathcal{B}_E$ always denotes the Borel σ-algebra of a topological space E.

Definition 1 (diffusion experiment). *Suppose $\Sigma \subset \Sigma(M_1, M_2)$ for some $M_1, M_2 > 0$. For any $T > 0$ let $\mathbb{E}(\Sigma, T)$ be the statistical experiment of observing the diffusion defined by (1) with $b \in \Sigma$, that is*

$$\mathbb{E}(\Sigma, T) = \big(C([0,T]; \mathbb{R}^d), \mathcal{B}_{C([0,T];\mathbb{R}^d)}, (\mathbf{P}_b^T)_{b\in\Sigma}\big).$$

For any function $b \in L^2(\mu_{b^\circ}; \mathbb{R}^d) = \{f : \mathbb{R}^d \to \mathbb{R}^d : \int |f|^2 \mu_{b^\circ} < \infty\}$ we denote by $\mathbf{Q}_{b,T}$ the Gaussian measure on $(C(\mathbb{R}^d; \mathbb{R}^d), \mathcal{B}_{C(\mathbb{R}^d;\mathbb{R}^d)})$ induced by the d-dimensional process Z satisfying

$$dZ(x) = b(x)\sqrt{\mu_{b^\circ}(x)}\, dx + T^{-1/2}\, dB(x), \qquad Z(\mathbf{0}) = \mathbf{0}, \qquad x \in \mathbb{R}^d, \quad (11)$$

where $B(x) = (B_1(x), \ldots, B_d(x))$ and $B_1(x), \ldots, B_d(x)$ are independent d-variate Brownian sheets, that is zero mean Gaussian processes with $\mathrm{Cov}(B_i(x), B_i(y)) = |R_x \cap R_y|$ where $R_x = \{u \in \mathbb{R}^d : u_i \in [0, x_i]\}$.

Definition 2 (Gaussian shift experiment). *For $\Sigma \subset L^2(\mu_{b^\circ}; \mathbb{R}^d)$ and $T > 0$ let $\mathbb{F}(\Sigma, T)$ be the Gaussian shift experiment (11) with $b \in \Sigma$, that is*

$$\mathbb{F}(\Sigma, T) = \big(C(\mathbb{R}^d; \mathbb{R}^d), \mathcal{B}_{C(\mathbb{R}^d;\mathbb{R}^d)}, (\mathbf{Q}_{b,T})_{b\in\Sigma}\big).$$

For any positive numbers ε, η and for any hypercube $A \subset \mathbb{R}^d$, we define the local neighbourhood of b°

$$\Sigma(b^\circ, \varepsilon, \eta, A) = \left\{ b \in \Sigma_\beta(L, M_1, M_2) : \begin{array}{c} |b(x) - b^\circ(x)| \le \varepsilon \mathbb{1}_A(x),\ x \in \mathbb{R}^d, \\ |\mu_b(x) - \mu_{b^\circ}(x)| \le \eta \mu_{b^\circ}(x),\ x \in A \end{array} \right\},$$

where $\mathbb{1}_A$ is the indicator function of the set A. We state the main local equivalence result, which will be proved in Section 6. The main ideas of the proof are explained in the next subsection. For the exact definition of statistical equivalence and the Le Cam distance Δ we refer to Le Cam and Yang [16].

Theorem 1. *If ε_T and η_T satisfy the conditions*

$$\lim_{T\to\infty} T^{-\beta}\varepsilon_T^{2-d} = \lim_{T\to\infty} T^{\frac{1}{4}+\frac{d-2}{8\beta}}\varepsilon_T(\log(T\varepsilon_T^{-1}))^{\mathbb{1}(d=2)} = \lim_{T\to\infty} T\eta_T\varepsilon_T^2 = 0,$$

then the diffusion model (1) *is asymptotically equivalent to the Gaussian shift model (11) over the parameter set $\Sigma_{0,T} = \Sigma(b^\circ, \varepsilon_T, \eta_T, A)$, that is*

$$\lim_{T\to\infty} \sup_{b^\circ\in\Sigma_\beta(L,M_1,M_2)} \Delta\big(\mathbb{E}(\Sigma_{0,T}, T), \mathbb{F}(\Sigma_{0,T}, T)\big) = 0.$$

Let us see for which Hölder regularity β on the drift an estimator can attain the local neighbourhood, that is $|\hat{b}_{h(T),T}(x) - b(x)| \le \varepsilon_T$ and $|\hat{\mu}_{h(T),T}(x) - \mu(x)| \le \eta_T$ hold with a probability tending to one (cf. Nussbaum [18] for this concept). By the rates obtained in Corollary 1, with a glance at Remark 7 and the condition in Theorem 1, this is the case if

$$\begin{aligned} -\beta - (2-d)\beta/(2\beta+d) &< 0, \\ 1/4 + (d-2)/(8\beta) - \beta/(2\beta+d) &< 0, \\ 1 - (\beta+1)/(2\beta+d) - 2\beta/(2\beta+d) &< 0. \end{aligned}$$

It turns out that the second condition is most binding and all three conditions are satisfied if $\beta > (d-1+\sqrt{2(d-1)^2-1})/2$. The critical regularity thus grows like $(1/2+1/\sqrt{2})d$ for $d\to\infty$. In dimension 2 we obtain the condition $\beta > 1$ as in the result by Carter [8] for Gaussian regression. Whether for Hölder classes of smaller regularity asymptotic equivalence fails, remains a challenging open problem.

3.2. Method of proof

The general idea of the proof of Theorem 1 consists in discretising (in space) the diffusion process such that the design regularisation technique we introduced in [10] is applicable in spirit, even though the local time does not exist.

Space discretisation. For any multi-index $\alpha \in \mathbb{N}^d$ set $\alpha! = \alpha_1! \cdot \ldots \cdot \alpha_d!$. Let us denote by $\{v_i\}_{i=1,\ldots,K}$ the elements of the set $\{v \in \mathbb{R}[x] : v(x) = x^\alpha \text{ with } |\alpha| \le \lfloor\beta\rfloor\}$ somehow enumerated: $v_i(x) = x_1^{\alpha_1(i)} \cdot \ldots \cdot x_d^{\alpha_d(i)} = x^{\alpha(i)}$. We assume that $A = [-a, a[^d$ is a hypercube and for some $h > 0$ with $a/h \in \mathbb{N}$ we denote by $\{a_m\}_{m=1,\ldots,M}$ the elements of the grid $(h\mathbb{Z}^d) \cap A$. We introduce the subcubes $\mathbf{C}_m = \prod_{j=1}^d [a_{mj}, a_{mj}+h[\subset A$, $m = 1,\ldots,M$, where a_{mj} is the jth coordinate of a_m. Let us define

$$\mathbf{v}(x) = \begin{pmatrix} v_1(x)/\alpha(1)! \\ \vdots \\ v_K(x)/\alpha(K)! \end{pmatrix}, \tag{12}$$

which gives rise to the definition $\bar{b}$ of the Taylor approximation for b

$$\bar{b}(x) = \sum_{i=1}^{K} D^{\alpha(i)} b(a_m)\mathbf{v}_i(x - a_m) \text{ for } x \in \mathbf{C}_m,\ m = 1, \dots, M$$

and $\bar{b}(x) = b^\circ(x)$ for $x \in \mathbb{R}^d \setminus A$ ($D^{\alpha(i)}$ is applied coordinate-wise). Using this notation, the Taylor formula can be written as

$$b(x) = \bar{b}(x) + \sum_{i:|\alpha(i)|=\lfloor\beta\rfloor} \Big(D^{\alpha(i)} b(\zeta) - D^{\alpha(i)} b(a_m)\Big) \frac{v_i(x - a_m)}{\alpha(i)!}, \quad x \in \mathbf{C}_m, \tag{13}$$

where $\zeta \in \mathbb{R}^d$ satisfies $|\zeta - a_m| \le |x - a_m|$. This implies that for $V \in \mathcal{H}(\beta + 1, L)$, the estimate $|b(x) - \bar{b}(x)| \lesssim h^\beta$ holds. We write

$$\vartheta(x) = b(x) - b^\circ(x), \quad \bar{\vartheta}(x) = \bar{b}(x) - \bar{b}^\circ(x) \text{ and } \boldsymbol{\theta}_j(x) = \begin{pmatrix} D^{\alpha(1)}\vartheta_j(x) \\ \vdots \\ D^{\alpha(K)}\vartheta_j(x) \end{pmatrix}$$

for $j = 1, \dots, d$ and we shall use equivalently θ and b for referring to the parameter in the local neighbourhood. The log-likelihood of the experiment defined via $\mathbf{P}_{\bar{b}}^T$ is given by (see Liptser and Shiryaev [17, p. 271, (7.62)])

$$\log \frac{d\mathbf{P}_{\bar{b}}^T}{d\mathbf{P}_{\bar{b}^\circ}^T}(X^T) = \sum_{m=1}^{M} \sum_{j=1}^{d} \Big[\boldsymbol{\theta}_j(a_m)^T \hat{\eta}_{mj}(T) - \frac{1}{2}\, \boldsymbol{\theta}_j(a_m)^T \hat{\mathcal{J}}_m(T)\, \boldsymbol{\theta}_j(a_m)\Big], \tag{14}$$

where

$$\begin{aligned} \hat{\eta}_{mj}(T) &= \int_0^T \mathbb{1}_{\mathbf{C}_m}(X_t)\mathbf{v}(X_t - a_m)\, dW_{t,j} \quad \in \mathbb{R}^K, \\ \hat{\mathcal{J}}_m(T) &= \int_0^T \mathbb{1}_{\mathbf{C}_m}(X_t)\mathbf{v}(X_t - a_m)\mathbf{v}(X_t - a_m)^T\, dt \quad \in \mathbb{R}^{K\times K}, \end{aligned} \tag{15}$$

and $W_{t,j}$ denotes the jth component of $W_t \in \mathbb{R}^d$.

Design modification. Due to the ergodicity of X the law of the log-likelihood (14) will for large T be well approximated by

$$\sum_{m=1}^{M} \sum_{j=1}^{d} \Big(\sqrt{T}\, \boldsymbol{\theta}_j(a_m)^T \eta_{mj} - \frac{T}{2}\, \boldsymbol{\theta}_j(a_m)^T \mathcal{J}_m \boldsymbol{\theta}_j(a_m)\Big) \tag{16}$$

where $\eta_{mj} \sim \mathcal{N}(0, \mathcal{J}_m)$ i.i.d. and

$$\mathcal{J}_m = \int_{\mathbf{C}_m} \mathbf{v}(x - a_m)\mathbf{v}(x - a_m)^T \mu_{b^\circ}(x)\, dx. \tag{17}$$

Since

$$\boldsymbol{\theta}_j(a_m)^T \mathcal{J}_m \boldsymbol{\theta}_j(a_m) = \int_{\mathbf{C}_m} (\bar{b}_j(x) - \bar{b}_j^\circ(x))^2 \mu_{b^\circ}(x)\,dx, \tag{18}$$

the process (16) (indexed by $\boldsymbol{\theta}$) has exactly the same law as the log-likelihood of the Gaussian shift

$$dZ(x) = \bar{b}(x)\sqrt{\mu_{b^\circ}(x)}\,dx + T^{-1/2}dB(x), \qquad Z(\mathbf{0}) = \mathbf{0}, \qquad x \in \mathbb{R}^d.$$

Under suitable assumptions on the smoothness of b, this last experiment is asymptotically equivalent to (11).

It remains to construct the random variables (η_{mj}) on some enlargement of the probability space $(C([0,T];\mathbb{R}^d), \mathcal{B}_{C([0,T];\mathbb{R}^d)}, \mathbf{P}_b^T)$ such that $T^{-1/2}\hat{\eta}_{mj}(T)$ and η_{mj} are close as random variables. We define the stopping time

$$\tau_m = \inf\left\{t \in [0,T] : \|\mathcal{J}_m^{-1/2}\hat{\mathcal{J}}_m(t)\mathcal{J}_m^{-1/2}\| \geq T\right\} \wedge T, \tag{19}$$

where the norm of a matrix A is given by $\|A\| = \sup_x(|Ax|/|x|)$.

Let $\boldsymbol{\varepsilon} = (\varepsilon_{mj})_{m,j}$ be a family of independent standard normal random vectors in $\mathbb{R}^K$, defined on an enlarged probability space such that $\boldsymbol{\varepsilon}$ and X are independent. We set

$$\eta_{mj} = \frac{1}{\sqrt{T}}\,\hat{\eta}_{mj}(\tau_m) + (\mathcal{J}_m - T^{-1}\hat{\mathcal{J}}_m(\tau_m))^{1/2}\varepsilon_{mj}.$$

By definition of τ_m the matrix $\mathcal{J}_m - T^{-1}\hat{\mathcal{J}}_m(\tau_m)$ is nonnegative definite and its square root is well defined.

Proposition 2. *Under the probability measure $\mathbf{P}_{b^\circ}^T$ the random vectors $(\eta_{mj})_{m,j} \subset \mathbb{R}^K$ are independent and each η_{mj} is centred Gaussian with covariance matrix $\mathcal{J}_m$.*

Proof. It suffices to show that for any sequence $(\lambda_{mj})_{m,j} \subset \mathbb{R}^K$ we have

$$\mathbf{E}\left[\exp\left\{\sum_{m,j}\lambda_{mj}^T\eta_{mj}\right\}\right] = \exp\left\{\frac{1}{2}\sum_{m,j}\lambda_{mj}^T\mathcal{J}_m\lambda_{mj}\right\},$$

where the expectation is taken with respect to X following the law $\mathbf{P}_{b^\circ}^T$ and ε_{mj} being i.i.d. standard normal in $\mathbb{R}^K$, independent of X.

The verification of this equality is very similar to the proof of Proposition 2.13 in Dalalyan and Reiß [10] and is omitted. □

4. Equivalence with heteroskedastic Gaussian regression

The Gaussian experiment in Theorem 1 depends on the centre b° of the neighbourhood via μ_{b°. This fact makes the passage from the local equivalence to a global equivalence difficult, especially, because even in the one-dimensional case there is no known variance stabilising transform for (11), cf. Dalalyan and Reiß [10].

We propose here a method of deriving an asymptotically equivalent experiment independent of b° without using the variance stabilising transform. The idea is to discretise the Gaussian shift experiment with a "step of discretisation" larger than $1/T$. This method has already been used in Brown and Zhao [7] for proving the asymptotic equivalence between regression models with random and deterministic designs.

We adopt the notation from Section 3.2. In addition, we introduce the $K \times K$-matrix $\mathbf{V} = \int_{[0,1]^d} \mathbf{v}(x)\mathbf{v}(x)^T\,dx$, where $\mathbf{v}(x)$ is defined by (12). Since $\mathbf{V}$ is strictly positive and symmetric, the matrix $\mathbf{V}^{-1/2}$ is well defined.

Definition 3 (heteroskedastic Gaussian regression). *Let Σ be a subset of $C^{\lfloor\beta\rfloor}(\mathbb{R}^d;\mathbb{R}^d)$. For any $T, h > 0$ we define $\mathbb{G}(\Sigma, h, T)$ as the experiment of observing*

$$Y_{im} = \begin{pmatrix} h^{|\alpha(1)|} D^{\alpha(1)} b_i \\ \vdots \\ h^{|\alpha(K)|} D^{\alpha(K)} b_i \end{pmatrix} (a_m) + \mathbf{V}^{-1/2} \frac{\xi_{im}}{\sqrt{T h^d \mu_b(a_m)}} \tag{20}$$

for $i = 1, \ldots, d$, $m = 1, \ldots, M$, where $(\xi_{im})_{i,m}$ is a family of independent standard Gaussian random vectors in $\mathbb{R}^K$ and $b \in \Sigma$.

Note that the observations in this experiment are chosen from $\mathbb{R}^{KMd}$ according to a Gaussian measure. Both the mean and the variance of this measure depend on the parameter b such that the experiment is heteroskedastic.

Theorem 2. *If the assumptions of Theorem 1 are fulfilled and $h = h_T$ satisfies*

$$\lim_{T\to\infty} T h_T^{2\beta} = \lim_{T\to\infty} T h_T^2 \varepsilon_T^2 = \lim_{T\to\infty} \eta_T^2 h_T^{-d} = 0,$$

then the diffusion experiments and the heteroskedastic Gaussian regression experiments are asymptotically equivalent, that is

$$\lim_{T\to\infty} \sup_{b^\circ \in \Sigma_\beta(L, M_1, M_2)} \Delta\big(\mathbb{E}(\Sigma_{0,T}, T), \mathbb{G}(\Sigma_{0,T}, h_T, T)\big) = 0.$$

Proof. Theorem 1 yields the asymptotic equivalence of the experiment $\mathbb{E}$ with the (translated) Gaussian shift experiment

$$d\widetilde{Z}(x) = (b - b^\circ)(x)\sqrt{\mu_{b^\circ}(x)}\,dx + T^{-1/2}dB(x), \quad x \in \mathbb{R}^d.$$

Let us introduce a new Gaussian shift:

$$d\widehat{Z}(x) = \sum_{m=1}^{M} \Big((\bar{b} - b^\circ)(x)\sqrt{\mu_{b^\circ}(a_m)}\Big)\mathbb{1}_{\mathbf{C}_m}(x)\,dx + T^{-1/2}dB(x), \quad x \in \mathbb{R}^d.$$

Since $|\nabla\mu_b(x)|$ and $|\mu_b(x)|$ are uniformly bounded, the difference between the drifts of $\widetilde{Z}$ and $\widehat{Z}$ can be estimated as follows:

$$\begin{aligned}&\big|(b - b^\circ)(x)\sqrt{\mu_{b^\circ}(x)} - (\bar{b} - b^\circ)(x)\sqrt{\mu_{b^\circ}(a_m)}\big|\\ &\quad\le \big|(b - \bar{b})(x)\sqrt{\mu_{b^\circ}(a_m)}\big| + \big|(b - b^\circ)(x)\big(\sqrt{\mu_{b^\circ}(x)} - \sqrt{\mu_{b^\circ}(a_m)}\big)\big|\\ &\quad\lesssim h^\beta + \varepsilon h \qquad \forall x \in \mathbf{C}_m.\end{aligned}$$

Therefore, the Hellinger distance between the measures induced by $\widetilde{Z}$ and $\widehat{Z}$ tends to zero as $T \to \infty$ (Strasser [23, Rem. 69.8.(2)]), provided that $T\varepsilon^2h^2 \to 0$ and $Th^{2\beta} \to 0$. The log-likelihood of the experiment given by $\widehat{Z}$ has exactly the same law as the log-likelihood of the Gaussian regression

$$Y_{im} = \begin{pmatrix} h^{|\alpha(1)|}D^{\alpha(1)}b_i \\ \vdots \\ h^{|\alpha(K)|}D^{\alpha(K)}b_i \end{pmatrix}(a_m) + \mathbf{V}^{-1/2}\frac{\xi_{im}}{\sqrt{Th^d\mu_{b^\circ}(a_m)}} \tag{21}$$

for $i = 1, \dots, d$; $m = 1, \dots, M$, where $(\xi_{im})_{i,m}$ is a family of independent standard Gaussian random vectors in $\mathbb{R}^K$ and $b \in \Sigma$. By Lemma 3 from Brown et *al.* [4] the square of the Hellinger distance between the measures induced by the observations (20) and (21), respectively, is up to a constant bounded by $\sum_{m=1}^{M}(\mu_b(a_m) - \mu_{b^\circ}(a_m))^2/\mu_{b^\circ}(a_m)^2 \lesssim M\eta_T^2$. Because of $Mh^d = |A|$ we infer $M \sim h^{-d}$ and the condition $h_T^{-d}\eta_T^2 \to 0$ as $T \to \infty$ implies that the Hellinger distance tends to zero uniformly in $b \in \Sigma_{0,T}$. Finally, the desired result follows by bounding the Le Cam distance between experiments by the supremum of the Hellinger distance between the corresponding measures, see e.g. Nussbaum [18, Eq. (12)]. □

Remark 8. The experiment given by (20) is more informative than the experiment generated by the observations $(\mathbf{e}_1^TY_{im})_{i,m}$, where $\mathbf{e}_1 = (1, 0, \dots, 0)^T \in \mathbb{R}^K$. If we enumerate $\{\alpha(i)\}_i$ so that $\alpha(1) = \mathbf{0} \in \mathbb{R}^d$ then $\widetilde{Y}_m := (\mathbf{e}_1^TY_{1m}, \dots, \mathbf{e}_1^TY_{dm})^T$ satisfies $\widetilde{Y}_m = b(a_m) + \epsilon_m/\sqrt{Th^d\mu_b(a_m)}$ with $\epsilon_m/\sqrt{(\mathbf{V}^{-1})_{11}} \sim \mathcal{N}(0, I_d)$ i.i.d. Therefore the diffusion experiment $\mathbb{E}(\Sigma_{0,T}, T)$ is asymptotically more informative than the regression experiment:

$$\widetilde{Y}_m = b(a_m) + \frac{\epsilon_m}{\sqrt{Th^d\mu_b(a_m)}}, \quad m = 1, \dots, M.$$

If we choose $h_T = T^{-\alpha}$, $\varepsilon_T = T^{-\beta/(2\beta+d)}$ and $\eta_T = T^{-(\beta+1)/(2\beta+d)}$ (in view of Corollary 1), the condition of Theorem 2 takes the form

$$\max\left(\frac{1}{\beta};\ \frac{d}{2\beta+d}\right) < 2\alpha < \frac{4(\beta+1)}{d(2\beta+d)}.$$

Such a value α exists if and only if

$$\beta > \max\left(\frac{d^2}{4} - 1;\ \frac{d-2+\sqrt{(d-2)^2+4d^2}}{4}\right).$$

For $d = 2$ this inequality reduces to $\beta > 1$. For $d \geq 4$ it is equivalent to $\beta > (d/2)^2 - 1$. Note also that the logarithmic factors in ε_T and η_T do not affect this bound on the minimal regularity.

As mentioned in the introduction, the result of Theorem 2 is new already in the one-dimensional case. When $d = 1$, using a $\sqrt{T}$-consistent estimator of μ_b (Kutoyants [15], § 4.2), the local neighbourhood can be attained as soon as $\beta > 1/2$. Taking $K = 1$ and using the globalisation method developed in [10], we obtain the global asymptotic equivalence of the diffusion experiment and the regression

$$Y_m = b(a_m) + \frac{\epsilon_m}{\sqrt{Th\mu_b(a_m)}}, \qquad m = 1, \ldots, M,$$

provided that $h = h_T = T^{-\alpha}$ with $(2\beta)^{-1} < \alpha < 1$ and the assumptions of [10, Thm. 3.5] are fulfilled.

5. Equivalence mapping

The result of Theorem 1 implies in particular that there exists a Markov kernel K from $(C([0,T];\mathbb{R}^d), \mathcal{B}_{C([0,T];\mathbb{R}^d)})$ to $(C(\mathbb{R}^d;\mathbb{R}^d), \mathcal{B}_{C(\mathbb{R}^d;\mathbb{R}^d)})$ such that

$$\lim_{T\to\infty} \sup_{b\in\Sigma_{0,T}} \|\mathbf{P}_b^T K - \mathbf{Q}_{b,T}\|_{TV} = 0,$$

where $\mathbf{P}_b^T K(A) = \int_{C([0,T];\mathbb{R}^d)} K(x,A)\mathbf{P}_b^T(dx)$ for $A \in \mathcal{B}_{C(\mathbb{R}^d;\mathbb{R}^d)}$ and $\|\cdot\|_{TV}$ denotes the total variation norm. The aim of this section is to construct this Markov kernel explicitly. The construction is divided into two steps. First, we give the Markov kernel from the diffusion experiment to a suitable multivariate Gaussian regression. Then we give the Markov kernel from the Gaussian regression to the Gaussian shift experiment. An explicit Markov kernel in the other direction is not known, but seems also less useful.

Assume that we have a path X^T of the diffusion process (1) at our disposal. In what follows we use the notation introduced in Section 3.2 with h verifying (27) below. For any $i = 1, \ldots, d$ we denote by $X_{t,i}$ the ith coordinate of X_t and define the randomisation

$$\Phi^{(1)}_{im}(X^T,\boldsymbol{\varepsilon}) = \frac{1}{T}\int_0^{\tau_m} \mathbb{1}_{\mathbf{C}_m}(X_t)\mathbf{v}(X_t-a_m)\,(dX_{t,i}-\bar{b}^\circ_i(X_t)\,dt) \\ +\frac{1}{\sqrt{T}}(\mathcal{J}_m-T^{-1}\hat{\mathcal{J}}_m(\tau_m))^{1/2}\varepsilon_{im}, \quad m=1,\dots,M,$$

where $\hat{\mathcal{J}}_m(t)$, $\mathcal{J}_m$ and τ_m are defined by (15), (17) and (19) and $\boldsymbol{\varepsilon}=(\varepsilon_{im})_{i,m}$ is a family of independent (and independent of X^T) standard Gaussian vectors in $\mathbb{R}^K$. As is easily checked, the random vector $\mathcal{J}_m^{-1}\Phi^{(1)}_{im}(X^T,\tilde{\boldsymbol{\varepsilon}})$ with $\tilde{\varepsilon}_{im}=(T\mathcal{J}_m-\hat{\mathcal{J}}_m(\tau_m))^{1/2}\boldsymbol{\theta}_i(a_m)+\varepsilon_{im}$ has the same law as the Gaussian regression

$$Y_{im}=\boldsymbol{\theta}_i(a_m)+(T\mathcal{J}_m)^{-1/2}\varepsilon_{im}. \tag{22}$$

We prove in Section 6.1 that the total variation between the laws of $\boldsymbol{\varepsilon}$ and $\tilde{\boldsymbol{\varepsilon}}$ tends to zero as $T\to\infty$. Consequently, if we denote by $K^{(1)}(x,\cdot)$ the law of $\{\mathcal{J}_m^{-1}\Phi^{(1)}_{im}(x,\boldsymbol{\varepsilon});\ i=1,\dots,d;\ m=1,\dots,M\}$, we obtain a Markov kernel realising the asymptotic equivalence between the diffusion (1) and the Gaussian regression (22).

For any $x\in\mathbf{C}_m$ and for any $i\in\{1,\dots,d\}$, we define the randomisation of the regression (22) by

$$\Phi^{(2)}_{i,x}\ (Y,\tilde{B}) = \int_{R(a_m,x)}\big(\bar{b}^\circ_i(u)+\mathbf{v}(u)^TY_{im}\big)\sqrt{\mu_{b^\circ}(u)}\,du \\ +\frac{1}{\sqrt{T}}\int_{R(a_m,x)}\sqrt{\mu_{b^\circ}(u)}\,d\tilde{B}_i(u) \\ -\frac{1}{\sqrt{T}}\Big(\int_{R(a_m,x)}\mathbf{v}(u)^T\mu_{b^\circ}(u)\,du\Big)\mathcal{J}_m^{-1}\Big(\int_{\mathbf{C}_m}\mathbf{v}(u)\sqrt{\mu_{b^\circ}(u)}\,d\tilde{B}_i(u)\Big), \tag{23}$$

where $R(a_m,x)=\prod_{i=1}^d[a_{mi},x_i[$, $\tilde{B}=(\tilde{B}_1,\dots,\tilde{B}_d)$ and $\tilde{B}_1,\dots,\tilde{B}_d$ are independent d-variate Brownian sheets independent of $(Y_{im})_{i,m}$. Let us show that $\Phi^{(2)}(\mathbf{y},\tilde{B})=(\Phi^{(2)}_{i,x}(\mathbf{y},\tilde{B});\,i\in\{1,\dots,d\},x\in A)$ is an equivalence mapping from the Gaussian regression model (22) to the Gaussian shift model (11).

For any $x\in\mathbf{C}_m$ and for any $i=1,\dots,d$ define the multivariate analogue of a Brownian bridge

$$V_i(x)=\int_{R(a_m,x)}\mathbf{v}(u)\sqrt{\mu_{b^\circ}(u)}\,d\tilde{B}_i(u) \\ -\Big(\int_{R(a_m,x)}\mathbf{v}(u)\mathbf{v}(u)^T\mu_{b^\circ}(u)\,du\Big)\mathcal{J}_m^{-1}\Big(\int_{\mathbf{C}_m}\mathbf{v}(u)\sqrt{\mu_{b^\circ}(u)}\,d\tilde{B}_i(u)\Big)$$

and set

$$\tilde{V}_i(x)=\Big(\int_{R(a_m,x)}\mathbf{v}(u)\mathbf{v}(u)^T\mu_{b^\circ}(u)\,du\Big)\,Y_{im}+T^{-1/2}V_i(x).$$

The process $\tilde{V}_i$ takes values in $\mathbb{R}^K$ and can be rewritten in the form $\tilde{V}_i(x) = \int_{R(a_m,x)} \mathbf{v}(u)(\bar{b}_i(u) - \bar{b}_i^\circ(u))\mu_{b^\circ}(u)\,du + T^{-1/2}\widehat{W}_i(x)$ where

$$\widehat{W}_i(x) = \Big(\int_{R(a_m,x)} \mathbf{v}(u)\mathbf{v}(u)^T \mu_{b^\circ}(u)\,du\Big)\mathcal{J}_m^{-1/2}\varepsilon_{im} + V_i(x).$$

By construction, the process $\widehat{W}_i$ is centred Gaussian with covariance matrix $\mathbf{E}[\widehat{W}_i(x)\widehat{W}_i(\bar{x})^T] = \int_{R(a_m,x)\cap R(a_m,\bar{x})} \mathbf{v}(u)\mathbf{v}(u)^T\mu_{b^\circ}(u)\,du$. Assuming that $v_1,\dots,v_K$ are enumerated in such a way that $v_1(u) \equiv 1$, one checks that $\widehat{B}_i(x) = \int_{R(a_m,x)} \mu_{b^\circ}(u)^{-1/2} d\widehat{W}_{i,1}(u)$ is a d-variate Brownian sheet, where $\widehat{W}_{i,1}$ is the first coordinate of $\widehat{W}_i$. Therefore, the randomisation

$$\Phi^{(2)}_{i,x}(Y,\tilde{B}) = \int_{R(a_m,x)} \bar{b}_i^\circ(u)\sqrt{\mu_{b^\circ}(u)}\,du + \int_{R(a_m,x)} \mu_{b^\circ}(u)^{-1/2} d\tilde{V}_{i,1}(u) \quad (24)$$

satisfies

$$d\Phi^{(2)}_{i,x} = \bar{b}_i(x)\sqrt{\mu_{b^\circ}(x)}\,dx + T^{-1/2}d\widehat{B}_i(x), \quad x \in \mathbf{C}_m, \quad i = 1,\dots d. \quad (25)$$

The total variation between the measures induced by (25) and (11) is up to a constant bounded by $\sqrt{T}h^\beta$, which tends to zero because of our choice of h and the assumptions of Theorem 1. Moreover, the d-variate Brownian sheets $\widehat{B}_1,\dots,\widehat{B}_d$ are independent. Simple algebra shows that the two definitions (24) and (23) coincide. Hence the law $K^{(2)}(\mathbf{y},\cdot)$ of $\Phi^{(2)}(\mathbf{y},\tilde{B})$ provides a Markov kernel from the Gaussian regression (22) to the Gaussian shift (11) realising the asymptotic equivalence.

6. Proof of Theorem 1

6.1. Main part

As we have seen in Section 3.2, the construction of the Gaussian experiment makes use of an i.i.d. family $\boldsymbol{\varepsilon} = (\varepsilon_{mj})_{m=1,\dots,M,\ j=1,\dots,d}$ of standard Gaussian vectors with values in $\mathbb{R}^K$. The canonical version of $\boldsymbol{\varepsilon}$ is defined on the measurable space $\big(\mathbb{R}^{KMd}, \mathcal{B}_{\mathbb{R}^{KMd}}\big)$. We prove the asymptotic equivalence by a suitable coupling, which consists in constructing probability measures $\tilde{\mathbf{P}}_b^T$ and $\tilde{\mathbf{Q}}_b^T$ on the product space

$$(\mathscr{E}, \mathcal{B}_{\mathscr{E}}) := \big(C([0,T],\mathbb{R}^d)\times\mathbb{R}^{KMd}, \mathcal{B}_{C([0,T],\mathbb{R}^d)}\otimes\mathcal{B}_{\mathbb{R}^{KMd}}\big)$$

such that

a) $\mathbb{E}(\Sigma_{0,T},T)$ is equivalent to $\tilde{\mathbb{E}}(\Sigma_{0,T},T) = \big(\mathscr{E},\mathcal{B}_{\mathscr{E}},(\tilde{\mathbf{P}}_b^T)_{b\in\Sigma_{0,T}}\big)$,
b) $\tilde{\mathbb{E}}(\Sigma_{0,T},T)$ and $\tilde{\mathbb{F}}(\Sigma_{0,T},T) = \big(\mathscr{E},\mathcal{B}_{\mathscr{E}},(\tilde{\mathbf{Q}}_b^T)_{b\in\Sigma_{0,T}}\big)$ are asymptotically equivalent,
c) $\mathbb{F}(\Sigma_{0,T},T)$ is asymptotically equivalent to $\tilde{\mathbb{F}}(\Sigma_{0,T},T)$.

a) Define $\tilde{\mathbf{P}}_b^T$ to be the measure induced by the pair $(X^T, \boldsymbol{\varepsilon})$, where X^T is given by (1) and $\boldsymbol{\varepsilon}$ is a standard Gaussian vector independent of X^T, that is $\tilde{\mathbf{P}}_b^T = \mathbf{P}_b^T \otimes \boldsymbol{\mathcal{N}}_{KMd}$ with $\boldsymbol{\mathcal{N}}_k$ denoting the standard normal law on $\mathbb{R}^k$. Then the equivalence $\mathbb{E} \sim \tilde{\mathbb{E}}$ follows from the equality in law of the respective likelihood processes, cf. Strasser [23, Cor. 25.9].

b) The measure $\tilde{\mathbf{Q}}_b^T$ is defined via

$$\tilde{\mathbf{Q}}_b^T(A \times B) = \int_{A\times B} e^{f_b(X^T,\boldsymbol{\varepsilon})}\, \mathbf{P}_{b^\circ}^T(dX^T)\, \boldsymbol{\mathcal{N}}_{KMd}(d\boldsymbol{\varepsilon})$$

for $A \in \mathcal{B}_{C([0,T],\mathbb{R}^d)}$ and $B \in \mathcal{B}_{\mathbb{R}^{KMd}}$ with

$$f_b(X^T, \boldsymbol{\varepsilon}) = \sum_{m=1}^{M} \sum_{j=1}^{d} \Big[\sqrt{T} \boldsymbol{\theta}_j(a_m)^T \eta_{mj}(X^T, \boldsymbol{\varepsilon}) - \frac{T}{2}\, \boldsymbol{\theta}_j(a_m)^T \mathcal{J}_m\, \boldsymbol{\theta}_j(a_m) \Big]$$

and

$$\begin{aligned}\eta_{mj}(X^T, \boldsymbol{\varepsilon}) =& \frac{1}{\sqrt{T}} \int_0^{\tau_m} \mathbb{1}_{\mathbf{C}_m}(X_t)\mathbf{v}(X_t - a_m)\,(dX_{t,j} - b_j^\circ(X_t)\,dt) \\ &+ (\mathcal{J}_m - T^{-1}\hat{\mathcal{J}}_m(\tau_m))^{1/2} \varepsilon_{mj}.\end{aligned}$$

Because of $f_{b^\circ}(X^T, \varepsilon) = 0$ these definitions yield $\tilde{\mathbf{Q}}_{b^\circ}^T = \tilde{\mathbf{P}}_{b^\circ}^T$ and therefore $\log\big(\frac{d\tilde{\mathbf{Q}}_b^T}{d\tilde{\mathbf{Q}}_{b^\circ}^T}\,(X^T, \boldsymbol{\varepsilon})\big) = f_b(X^T, \boldsymbol{\varepsilon})$. Proposition 2 combined with the classical formula of the characteristic function of a Gaussian vector implies that $\tilde{\mathbf{Q}}_b^T$ is a probability measure.

To prove the asymptotic equivalence of $\tilde{\mathbb{E}}$ and $\tilde{\mathbb{F}}$, it suffices to show that the Kullback-Leibler divergence between the measures $\tilde{\mathbf{P}}_b^T$ and $\tilde{\mathbf{Q}}_b^T$ tends to zero uniformly in $b \in \Sigma_{0,T}$ (see the proof of Thm. 2.16 in [10]). The Fubini theorem yields

$$\begin{aligned}KL(\tilde{\mathbf{P}}_b^T, \tilde{\mathbf{Q}}_b^T) &= \int \log\Big(\frac{d\tilde{\mathbf{P}}_b^T}{d\tilde{\mathbf{Q}}_b^T}(X^T, \boldsymbol{\varepsilon})\Big) \mathbf{P}_b^T(dX^T) \boldsymbol{\mathcal{N}}_{KMd}(d\boldsymbol{\varepsilon}) \\ &= \mathbf{E}_b\Big[\log\Big(\frac{d\mathbf{P}_b^T}{d\mathbf{P}_{b^\circ}^T}(X^T)\Big) - \int f_b(X^T, \boldsymbol{\varepsilon})\, \boldsymbol{\mathcal{N}}_{KMd}(d\boldsymbol{\varepsilon})\Big].\end{aligned}$$

The Girsanov formula (Liptser and Shiryaev [17]) and the fact that the expectation of the stochastic integral is zero give

$$\begin{aligned}&\mathbf{E}_b\Big[\log\Big(\frac{d\mathbf{P}_b^T}{d\mathbf{P}_{b^\circ}^T}(X^T)\Big)\Big] = \mathbf{E}_b\Big[\log\Big(\frac{\mu_b(X_0)}{\mu_{b^\circ}(X_0)}\Big)\Big] + \frac{1}{2}\mathbf{E}_b\Big[\int_0^T |\vartheta(X_t)|^2\, dt\Big] \\ &= \mathbf{E}_b\Big[\log\Big(\frac{\mu_b(X_0)}{\mu_{b^\circ}(X_0)}\Big)\Big] + \frac{T}{2}\int_A \big|\vartheta(x) - \bar{\vartheta}(x)\big|^2 \mu_b(x)\, dx \\ &\quad + \frac{T}{2}\int_A |\bar{\vartheta}(x)|^2 \mu_b(x)\, dx + T\int_A \bar{\vartheta}(x)^T \big(\vartheta(x) - \bar{\vartheta}(x)\big)\, \mu_b(x)\, dx.\end{aligned}$$

Similarly, we find

$$\begin{aligned}\mathbf{E}_b\Big[\int f_b(X^T,\boldsymbol{\varepsilon})\,\boldsymbol{\mathcal{N}}_{KMd}(d\boldsymbol{\varepsilon})\Big] &= \sum_{m=1}^{M}\sum_{j=1}^{d}\Big(-\frac{T}{2}\,\boldsymbol{\theta}_j(a_m)^T\mathcal{J}_m\,\boldsymbol{\theta}_j(a_m)\\ &\qquad+\mathbf{E}_b\Big[\boldsymbol{\theta}_j(a_m)^T\int_0^{\tau_m}\mathbb{1}_{\mathbf{C}_m}(X_t)\mathbf{v}(X_t-a_m)\,\vartheta_j(X_t)\,dt\Big]\Big)\\ &= -\frac{T}{2}\int_A|\bar\vartheta(x)|^2\mu_{b^\circ}(x)\,dx+\sum_{m=1}^{M}\mathbf{E}_b\Big[\int_0^{\tau_m}\mathbb{1}_{\mathbf{C}_m}(X_t)|\bar\vartheta(X_t)|^2\,dt\Big]\\ &\qquad+\sum_{m=1}^{M}\mathbf{E}_b\Big[\int_0^{\tau_m}\mathbb{1}_{\mathbf{C}_m}(X_t)\bar\vartheta(X_t)^T(\vartheta(X_t)-\bar\vartheta(X_t))\,dt\Big].\end{aligned}$$

Using for $f(x)=|\bar\vartheta(x)|^2$ and $f(x)=\bar\vartheta(x)^T\big(\vartheta(x)-\bar\vartheta(x)\big)$ the general identity

$$T\int_A f(x)\mu_b(x)\,dx=\sum_{m=1}^{M}\mathbf{E}_b\Big[\int_0^T\mathbb{1}_{\mathbf{C}_m}(X_t)\,f(X_t)\,dt\Big],$$

we obtain $\mathrm{KL}(\tilde{\mathbf{P}}_b^T,\tilde{\mathbf{Q}}_b^T)=\sum_{i=1}^5\mathcal{T}_i(\vartheta)$ with

$$\begin{aligned}\mathcal{T}_1(\vartheta) &= \mathbf{E}_b\big[\log\mu_b(X_0)-\log\mu_{b^\circ}(X_0)\big],\\ \mathcal{T}_2(\vartheta) &= \frac{T}{2}\int_A|\bar\vartheta(x)|^2\big(\mu_{b^\circ}(x)-\mu_b(x)\big)\,dx,\\ \mathcal{T}_3(\vartheta) &= \sum_{m=1}^{M}\mathbf{E}_b\Big[\int_{\tau_m}^T|\bar\vartheta(X_t)|^2\mathbb{1}_{\mathbf{C}_m}(X_t)\,dt\Big],\\ \mathcal{T}_4(\vartheta) &= \frac{T}{2}\int_A\big|\vartheta(x)-\bar\vartheta(x)\big|^2\mu_b(x)\,dx,\\ \mathcal{T}_5(\vartheta) &= \sum_{m=1}^{M}\mathbf{E}_b\Big[\int_{\tau_m}^T\mathbb{1}_{\mathbf{C}_m}(X_t)\,\bar\vartheta(X_t)^T(\vartheta(X_t)-\bar\vartheta(X_t))\,dt\Big].\end{aligned}$$

The Cauchy-Schwarz inequality implies that $\mathcal{T}_5(\vartheta)\le\mathcal{T}_3(\vartheta)+\mathcal{T}_4(\vartheta)$. The explicit form of the invariant density μ_b implies that $\sup_\vartheta\mathcal{T}_1(\vartheta)\lesssim\varepsilon$. The Hölder assumption implies that $\sup_x|\bar\vartheta(x)-\vartheta(x)|\lesssim h^\beta$ and we infer

$$\sup_\vartheta\mathcal{T}_2(\vartheta)\lesssim T(h^{2\beta}+\varepsilon^2)\eta,\qquad\sup_\vartheta\mathcal{T}_4(\vartheta)\lesssim Th^{2\beta}.$$

In Section 6.2 below we prove that

$$\mathcal{T}_3(\vartheta)\lesssim(T\eta+\psi_d(h^d)\sqrt{T})\,\|\bar\vartheta\|_\infty^2 \tag{26}$$

holds if $h=h_T$ tends to zero for $T\to\infty$. Hence, we obtain

$$KL(\tilde{\mathbf{P}}_b^T,\tilde{\mathbf{Q}}_b^T)\lesssim\varepsilon+Th^{2\beta}+T(\varepsilon^2+h^{2\beta})\eta+\psi_d(h^d)\sqrt{T}(\varepsilon^2+h^{2\beta}).$$

Consequently, the rate-optimal choice of h is

$$h=h_T=(\varepsilon^4T^{-1})^{1/(4\beta+d-2)}, \tag{27}$$

provided that $h^{2\beta} = o(\varepsilon^2)$, so that

$$KL(\tilde{\mathbf{P}}_b^T, \tilde{\mathbf{Q}}_b^T) \lesssim \varepsilon + (\varepsilon^2 T^{\frac{1}{2}+\frac{d-2}{4\beta}})^{4\beta/(4\beta+d-2)} (\log(T\varepsilon^{-1}))^{2\mathbb{1}(d=2)} + T\varepsilon^2\eta,$$

given $\varepsilon^{d-2}T^{\beta} \to \infty$. Under the assumptions of the theorem we thus conclude that $\tilde{\mathbb{E}}$ and $\tilde{\mathbb{F}}$ are asymptotically equivalent.

c) It remains to verify that the statistical experiment $\mathbb{F}$ defined via $\mathbf{Q}_b^T$ is asymptotically equivalent to the experiment $\tilde{\mathbb{F}}$ defined via $\tilde{\mathbf{Q}}_b^T$. We have already seen that

$$\log\left(\frac{d\tilde{\mathbf{Q}}_b^T}{d\tilde{\mathbf{Q}}_{b^\circ}^T}\right) = \sum_{m,j}\left[\sqrt{T}\boldsymbol{\theta}_j(a_m)^T\eta_{mj} - \frac{T}{2}\,\boldsymbol{\theta}_j(a_m)^T\mathcal{J}_m\,\boldsymbol{\theta}_j(a_m)\right].$$

Recall that according to Proposition 2 the random vectors $(\eta_{mj})_{m,j}$ are independent Gaussian with covariance matrix $\mathcal{J}_m$. Therefore, the law of the log-likelihood process $\big(d\tilde{\mathbf{Q}}_b^T/d\tilde{\mathbf{Q}}_{b^\circ}^T\big)_{b\in\Sigma_0}$ coincides with the law of the process $\big(d\tilde{\mathbf{Q}}_{\bar{b}}^T/d\tilde{\mathbf{Q}}_{\bar{b}^\circ}^T\big)_{b\in\Sigma_0}$. This gives the equivalence of the experiments $\tilde{\mathbb{F}}$ and $\widehat{\mathbb{F}}$, where the latter experiment is defined by the observation

$$dZ(x) = \bar{b}(x)\sqrt{\mu_{b^\circ}(x)}\,dx + T^{-1/2}\,dB(x), \qquad Z(\mathbf{0}) = \mathbf{0}, \qquad x\in\mathbb{R}^d. \quad (28)$$

To conclude, we remark that the Kullback-Leibler divergence between the Gaussian experiments $\mathbb{F}$ and $\widehat{\mathbb{F}}$ is bounded by $T\int_{\mathbb{R}^d}(\bar{b}-b)^2\mu_{b^\circ} \le Th_T^{2\beta}$ and in view of (27) tends to zero for $T\to\infty$. $\square$

6.2. Evaluation of $\mathcal{T}_3$

We start by sketching how the estimate could be reduced to a purely analytical problem, using

$$\begin{aligned}\mathcal{T}_3(\vartheta) &\le \|\bar{b}-\bar{b}_0\|_\infty^2 \sum_m \mathbf{E}_b\Big[\int_{\tau_m}^T \mathbb{1}_{\mathbf{C}_m}(X_t)\,dt\Big] \qquad (29)\\ &\le \|\bar{b}-\bar{b}_0\|_\infty^2\Big(\sup_m \mathbf{E}_b[T-\tau_m] + \sum_m\Big(\mathbf{E}_b\Big[\int_{\tau_m}^T(\mathbb{1}_{\mathbf{C}_m}(X_t)-\mathbf{P}_b(\mathbf{C}_m))\,dt\Big]\Big).\end{aligned}$$

If f is a function in the domain of the generator L_b of the semigroup $(P_{b,t})_{t\ge0}$ with $L_bf = \mathbb{1}_{\mathbf{C}_m}(X_t) - \mathbf{P}_b(\mathbf{C}_m)$, then Dynkin's formula and the fact that $\mathbb{1}_{\mathbf{C}_m}(X_t) - \mathbf{P}_b(\mathbf{C}_m)$ is centred yield

$$\mathbf{E}_b\Big[\int_{\tau_m}^T(\mathbb{1}_{\mathbf{C}_m}(X_t)-\mathbf{P}_b(\mathbf{C}_m))\,dt\Big] = \mathbf{E}_b[f(X_{\tau_m})] \le \sup_{x\in\mathbf{C}_m} f(x).$$

Unfortunately, a suitably tight supremum norm estimate for $f = L_b^{-1}(\mathbb{1}_{\mathbf{C}_m} - \mathbf{P}_b(\mathbf{C}_m))$ could not be found in the literature.

We therefore proceed differently and make use of the mixing properties of X. Fix some $\Delta = \Delta(T) > 0$. Since for $\tau_m > T - \Delta$ the integral over $[\tau_m, T]$ is smaller than the integral over $[T - \Delta, T]$, we have

$$\mathbf{E}_b\Big[\int_{\tau_m}^{T} \mathbb{1}_{\{X_t \in \mathbf{C}_m\}}\, dt\Big] \le \Delta\mu_b(C_m) + \mathbf{E}_b\Big[\mathbb{1}_{\{\tau_m \le T-\Delta\}} \int_{\tau_m}^{T} \mathbb{1}_{\{X_t \in \mathbf{C}_m\}}\, dt\Big]. \tag{30}$$

Lemma 1. *Under the assumptions of Proposition 1 we obtain*

$$\mathbf{E}_b\Big[\mathbb{1}_{\{\tau_m \le T-\Delta\}} \int_{\tau_m}^{\tau_m+\Delta} \mathbb{1}_{\{X_t \in \mathbf{C}_m\}}\, dt\Big] \lesssim \Delta\mu_b(\mathbf{C}_m) + h^{\frac{d}{2}}\psi_d(h^d)\sqrt{T\mu_b(\mathbf{C}_m)}.$$

Proof. Because of $[\tau_m, \tau_m + \Delta] \subset [(i-1)\Delta, (i+1)\Delta]$ for some $1 \le i \le T/\Delta$ we get

$$\int_{\tau_m}^{\tau_m+\Delta} \mathbb{1}_{\mathbf{C}_m}(X_s)\, ds \le \max_{i=1,\ldots,[T/\Delta]} \int_{(i-1)\Delta}^{(i+1)\Delta} \mathbb{1}_{\mathbf{C}_m}(X_s)\, ds.$$

Set $U_i = \int_{(i-1)\Delta}^{(i+1)\Delta} \mathbb{1}_{\mathbf{C}_m}(X_s)\, ds - 2\Delta\mu_b(\mathbf{C}_m)$. By separating the bias from the stochastic term, we find

$$\int_{\tau_m}^{\tau_m+\Delta} \mathbb{1}_{\mathbf{C}_m}(X_s)\, ds \le 2\Delta\mu_b(\mathbf{C}_m) + \max_{i=1,\ldots,[T/\Delta]} |U_i|,$$

and by the Cauchy-Schwarz inequality

$$\mathbf{E}_b\big[\max_i |U_i|\big] \le \Big(\sum_{i=1}^{\lfloor T/\Delta\rfloor} \mathbf{E}_b(U_i^2)\Big)^{\frac{1}{2}} = \lfloor T/\Delta\rfloor^{1/2}\, \mathrm{Var}\left(\int_0^{2\Delta} \mathbb{1}_{\mathbf{C}_m}(X_s)\, ds\right)^{\frac{1}{2}}.$$

We conclude by an application of Proposition 1. □

Lemma 2. *If Assumption 1 is satisfied, then*

$$\mathbf{E}_b\Big[\mathbb{1}_{\{\tau_m \le T-\Delta\}} \int_{\tau_m+\Delta}^{T} \mathbb{1}_{\{X_t \in \mathbf{C}_m\}}\, dt\Big] \le \mu_b(\mathbf{C}_m) \int_0^{T-\Delta} \mathbf{P}_b^T(\tau_m \le t)\, dt + Te^{-\Delta\rho}\sqrt{\mu_b(\mathbf{C}_m)}.$$

Proof. We have

$$\begin{aligned}\mathbf{E}_b\Big[\mathbb{1}_{\{\tau_m \le T-\Delta\}} \int_{\tau_m+\Delta}^{T} \mathbb{1}_{\{X_t \in \mathbf{C}_m\}}\, dt\Big] &= \mathbf{E}_b\Big[\int_{\Delta}^{T} \mathbb{1}_{\{X_t \in \mathbf{C}_m\}}\, \mathbb{1}_{\{\tau_m \le t-\Delta\}}\, dt\Big] \\ &= \mu_b(\mathbf{C}_m) \int_{\Delta}^{T} \mathbf{P}_b^T(\tau_m \le t - \Delta)\, dt \\ &\quad + \int_{\Delta}^{T} \mathbf{E}_b\big[\big(\mathbb{1}_{\mathbf{C}_m}(X_t) - \mu_b(\mathbf{C}_m)\big)\, \mathbb{1}_{\{\tau_m \le t-\Delta\}}\big]\, dt.\end{aligned}$$

Using the Markov property of the process (X_t) and the spectral gap inequality from Assumption 1, we infer that

$$\begin{aligned}
&\mathbf{E}_b\big[\big(\mathbb{1}_{\mathbf{C}_m}(X_t)-\mu_b(\mathbf{C}_m)\big)\,\mathbb{1}_{\{\tau_m\le t-\Delta\}}\big]\\
&\qquad=\mathbf{E}_b\big[P_{b,\Delta}(\mathbb{1}_{\mathbf{C}_m}-\mu_b(\mathbf{C}_m))(X_{t-\Delta})\,\mathbb{1}_{\{\tau_m\le t-\Delta\}}\big]\\
&\qquad\le\sqrt{\mathbf{E}_b\big[\big(P_{b,\Delta}(\mathbb{1}_{\mathbf{C}_m}-\mu_b(\mathbf{C}_m))(X_{t-\Delta})\big)^2\big]}\\
&\qquad=\|P_{b,\Delta}\mathbb{1}_{\mathbf{C}_m}-\mu_b(\mathbf{C}_m)\|_{\mu_b}\le e^{-\Delta\rho}\sqrt{\mu_b(\mathbf{C}_m)}.
\end{aligned}$$

This inequality completes the proof of the lemma. □

Lemma 3. *We have uniformly over $m=1,\dots,M$:*

$$\mathbf{P}_b(\tau_m\le t)\lesssim\frac{t^2\eta^2+t\psi_d^2(h^d)}{(T-t)^2}.$$

Proof. Note that $M_t:=\mathcal{J}_m^{-1/2}\hat\eta_{mj}(t)\in\mathbb{R}^K$ is a martingale with quadratic variation matrix $\langle M\rangle_t=\mathcal{J}_m^{-1/2}\hat{\mathcal{J}}_m(t)\mathcal{J}_m^{-1/2}$. We obtain that $\mathbf{E}_b[\langle M\rangle_t]=tI_K$ with the $K\times K$-unit matrix I_K and

$$\begin{aligned}
\mathbf{P}_b(\tau_m\le t)&=\mathbf{P}_b(\|\langle M\rangle_t\|\ge T)=\mathbf{P}_b(\|\langle M\rangle_t-tI_K\|\ge T-t)\\
&\le\frac{\mathbf{E}_b[\|\langle M\rangle_t-tI_K\|^2]}{(T-t)^2}\ .
\end{aligned}$$

Let $J_h\in\mathbb{R}^{K\times K}$ be the diagonal matrix with $J_{h,ii}=h^{|\alpha(i)|}$, $i=1,\dots,K$, then

$$\begin{aligned}
\|\langle M\rangle_t-tI_K\|&=\|\mathcal{J}_m^{-1/2}(\hat{\mathcal{J}}_m(t)-t\mathcal{J}_m)\mathcal{J}_m^{-1/2}\|\\
&\le\|\mathcal{J}_m^{-1/2}J_h\|^2\|J_h^{-1}(\hat{\mathcal{J}}_m(t)-t\mathcal{J}_m)J_h^{-1}\|.
\end{aligned}$$

Simple algebra shows that $\|\mathcal{J}_m^{-1/2}J_h\|^2=\|(J_h^{-1}\mathcal{J}_mJ_h^{-1})^{-1}\|$, $J_h^{-1}=J_{h^{-1}}$ and

$$J_h^{-1}\mathcal{J}_mJ_h^{-1}=h^d\int_{[0,1]^d}\mathbf{v}(u)\mathbf{v}(u)^T\mu_{b^\circ}(a_m+uh)\,du.$$

This matrix is strictly positive definite and $\|h^{-d}\mu_{b^\circ}(a_m)^{-1}J_h^{-1}\mathcal{J}_mJ_h^{-1}-\mathbf{V}\|$ tends to zero as $h\to0$. Hence, by the continuity of the matrix inversion we obtain for h small enough

$$\|h^d\mu_{b^\circ}(a_m)J_h\mathcal{J}_m^{-1}J_h\|\le2\|\mathbf{V}^{-1}\|.$$

We conclude that $\|\mathcal{J}_m^{-1/2}J_h\|^2\lesssim\mu_{b^\circ}(\mathbf{C}_m)^{-1}$. Set now $H_t=J_h^{-1}(\hat{\mathcal{J}}_m(t)-t\mathcal{J}_m)J_h^{-1}$. It is easily checked that

$$\begin{aligned}
H_t=&\int_0^t\mathbb{1}_{\mathbf{C}_m}(X_s)\mathbf{v}\Big(\frac{X_s-a_m}{h}\Big)\mathbf{v}\Big(\frac{X_s-a_m}{h}\Big)^Tds\\
&-t\int_{\mathbf{C}_m}\mathbf{v}\Big(\frac{x-a_m}{h}\Big)\mathbf{v}\Big(\frac{x-a_m}{h}\Big)^T\mu_{b^\circ}(x)\,dx.
\end{aligned}$$

Each entry $H_{t,ij}$ can be written as $\int_0^t f(X_s)\,ds - t\int_{\mathbf{C}_m} f(x)\mu_{b^\circ}(x)\,dx$, where f is a function bounded by 1 and supported by $\mathbf{C}_m$. Thus, a bias-variance decomposition combined with Proposition 1 yields

$$\mathbf{E}_b[H_{t,ij}^2] \lesssim t^2 \Big(\int_{\mathbf{C}_m} |\mu_b(x) - \mu_{b^\circ}(x)|\,dx \Big)^2 + t h^d \psi_d^2(h^d)\mu_b(\mathbf{C}_m).$$

Since in view of Remark 6 $\mu_b(\mathbf{C}_m)$ and $\mu_{b^\circ}(\mathbf{C}_m)$ are both of order h^d and all norms in $\mathbb{R}^{K\times K}$ are equivalent, we arrive at the desired estimate. □

Using the last lemma we obtain

$$\begin{aligned}\int_0^{T-\Delta} \mathbf{P}_b(\tau_m \le t)\,dt &\lesssim \int_0^T \min\Big(1, \frac{t^2\eta^2}{(T-t)^2} + \frac{\psi_d^2(h^d)t}{(T-t)^2}\Big)\,dt \\ &\le \int_0^T \min\Big(1, \frac{t^2\eta^2}{(T-t)^2}\Big)\,dt + \int_0^T \min\Big(1, \frac{\psi_d^2(h^d)t}{(T-t)^2}\Big)\,dt.\end{aligned}$$

Setting $c_T = T^{-1/2}\psi_d(h^d)$, we get

$$\begin{aligned}\int_0^T \min\Big(1, \frac{\psi_d^2(h^d)t}{(T-t)^2}\Big)\,dt &= T\int_0^1 \min(1, c_T^2(1-v)v^{-2})\,dv \\ &\le T\int_0^{c_T} 1\,dv + T\int_{c_T}^\infty c_T^2 v^{-2}\,dv \\ &= 2Tc_T = 2T^{1/2}\psi_d(h^d).\end{aligned}$$

In the same way we obtain $\int_0^T \min\big(1, t^2\eta^2/(T-t)^2\big)\,dt \le 2T\eta$. Substituting all estimates into (30) and (29), we obtain

$$\mathcal{T}_3(\vartheta) \lesssim \|\bar b - \bar b^\circ\|_\infty^2 \big(\Delta + T\eta + \psi_d(h^d)\sqrt{T} + Th^{-d/2}e^{-\Delta\rho}\big).$$

Thus choosing $\Delta(T) = \psi_d(h^d)\sqrt{T}$ we get

$$\mathcal{T}_3(\vartheta) \lesssim \|\bar b - \bar b^\circ\|_\infty^2\,(T\eta + \psi_d(h^d)\sqrt{T}),$$

provided that $h = h(T)$ tends to zero as $T \to \infty$.

References

1. Aït-Sahalia, Y.: Closed-form likelihood expansions for multivariate diffusions. Preprint (2004), available under `http://www.princeton.edu/~yacine/research.htm`.
2. Bandi, F. M., Moloche, G.: On the functional estimation of multivariate diffusion processes. Preprint (2001), available under `http://gsbwww.uchicago.edu/fac/federico.bandi/research`.
3. Bhattacharya, R. N.: Criteria for recurrence and existence of invariant measures for multidimensional diffusions. Ann. Probab. **6**(4), 541–553 (1978)

4. Brown, L. D., Cai, T. T., Low, M. G. and Zhang, C.-H.: Asymptotic equivalence theory for nonparametric regression with random design. Dedicated to the memory of Lucien Le Cam. Ann. Statist. **30**(3), 688–707 (2002)
5. Brown, L. D., Low, M. G.: Asymptotic equivalence of nonparametric regression and white noise. Ann. Stat. **24**(6), 2384–2398 (1996)
6. Brown, L. D., Zhang, C.: Asymptotic nonequivalence of nonparametric experiments when the smoothness index is 1/2. Ann. Stat. **26**(1), 279–287 (1998)
7. Brown, L. D. and Zhao, L. Direct asymptotic equivalence of nonparametric regression and the infinite dimensional location problem. Preprint (2003), available under `http://www-stat.wharton.upenn.edu/~lzhao/papers`.
8. Carter, A. V.: A continuous Gaussian process approximation to a nonparametric regression in two dimensions. Preprint (2004), available under `http://www.pstat.ucsb.edu/faculty/carter`.
9. Chen, M.-F., Wang, F.-Y.: Estimation of spectral gap for elliptic operators. Trans. Amer. Math. Soc. **349**(3), 1239–1267 (1997)
10. Dalalyan, A. S., Reiß, M.: Asymptotic statistical equivalence for scalar ergodic diffusions. Probab. Theory Rel. Fields, to appear.
11. Delattre, S., Hoffmann, M.: Asymptotic equivalence for a null recurrent diffusion, Bernoulli **8**(2), 139–174 (2002)
12. Efromovich, S.: Nonparametric curve estimation. Methods, theory, and applications. Springer Series in Statistics. Springer-Verlag, New York, 1999
13. Giné, E., Koltchinskii, V., Zinn, J.: Weighted uniform consistency of kernel density estimators. Ann. Prob. **32**(3b), 2570–2605 (2004)
14. Korostelev, A., Nussbaum, M.: The asymptotic minimax constant for sup-norm loss in nonparametric density estimation. Bernoulli **5**(6), no. 6, 1099–1118 (1999)
15. Kutoyants, Y. A.: Statistical inference for ergodic diffusion processes. Springer Series in Statistics, London, 2004
16. Le Cam, L., Yang, G. L.: Asymptotics in statistics. Some basic concepts. Springer Series in Statistics. Springer-Verlag, New York, 1990
17. Liptser, R. S., Shiryaev, A. N.: Statistics of random processes. 1: General theory. 2nd edition, Applications of Mathematics 5. Springer, Berlin, 2001
18. Nussbaum, M.: Asymptotic equivalence of density estimation and Gaussian white noise. Ann. Stat. **24**(6), 2399–2430 (1996)
19. Qian, Z., Russo, F., Zheng, W.: Comparison theorem and estimates for transition probability densities of diffusion processes. Probab. Theory Related Fields **127**(3), 388–406 (2003)
20. Qian, Z., Zheng, W.: A representation formula for transition probability densities of diffusions and applications. Stochastic Process. Appl. **111**(1), 57–76 (2004)
21. Revuz, D.,Yor, M.: Continuous martingales and Brownian motion. 3rd ed. Graduate Texts in Mathematics 293, Springer, Berlin, 1999
22. Rogers, L. C. G., Williams, D.: Diffusions, Markov processes, and martingales. Vol. 2. Itô calculus. Wiley Series in Probability and Mathematical Statistics. New York, 1987
23. Strasser, H.: Mathematical theory of statistics. Statistical experiments and asymptotic decision theory. De Gruyter Studies in Mathematics, 7. Walter de Gruyter, Berlin, 1985