

ADAPTIVE INVARIANT DENSITY ESTIMATION FOR ERGODIC DIFFUSIONS OVER ANISOTROPIC CLASSES

BY CLAUDIA STRAUCH¹

Universität Mannheim

Consider some multivariate diffusion process $\mathbf{X} = (X_t)_{t \geq 0}$ with unique invariant probability measure and associated invariant density ρ , and assume that a continuous record of observations $X^T = (X_t)_{0 \leq t \leq T}$ of \mathbf{X} is available. Recent results on functional inequalities for symmetric Markov semi-groups are used in the statistical analysis of kernel estimators $\hat{\rho}_T = \hat{\rho}_T(X^T)$ of ρ . For the basic problem of estimation with respect to sup-norm risk under anisotropic Hölder smoothness constraints, the proposed approach yields an adaptive estimator which converges at a substantially faster rate than in standard multivariate density estimation from i.i.d. observations.

1. Introduction. In many areas of applied mathematics, diffusion phenomena arise from a Markovian stochastic modelling and as a solution of SDEs or PDEs. Their investigation thus concerns different mathematical branches, starting from probability theory over functional analysis up to differential geometry for operators and the study of processes on manifolds. In all of these fields, research interest has constantly grown over the past years and decades with vivid interest in questions such as existence and regularity of solutions of SDEs, regularity and smoothing properties of differential operators, convergence to equilibrium and so on. The study of the statistical properties of diffusion models has emerged since such models are heavily used in applications from finance. Practical concerns raise new questions to date. As a matter of fact, in comparison to “classical” statistical models, a number of phenomena and peculiarities arises in this comparatively young branch of mathematical statistics. One such peculiarity concerns the dependence of statistical features on the observation scheme such that it is of interest to study basic questions—such as the estimation of the characteristics of the diffusion process—in different observation sceneries. By way of example, an original question (which has no counterpart in “classical” statistics) concerns the nonparametric estimation of the drift and diffusion coefficient of a diffusion process from low-frequency observations. In the scalar setup, this problem was solved in [9]. Nonparametric estimation of the characteristics of a reflected scalar diffusion on a

Received March 2017; revised October 2017.

¹Supported by the Deutsche Forschungsgemeinschaft (DFG) through the Research Training Group RTG 1953.

MSC2010 subject classifications. 62M05, 62G07, 62G20.

Key words and phrases. Ergodic diffusion, anisotropic density estimation, adaptation.

one-dimensional interval when the process is observed at random times was studied recently in [6].

From a theoretical point of view, it is however also of substantial interest to work under the assumption that a continuous record of observations of the diffusion is available. In this framework, there exist different examples where the smoothing properties of transition operators of Markovian semigroups allow to derive results which reflect somehow increased regularity in statistical diffusion models. For instance, it belongs to the folklore of statistics for stochastic processes that the invariant density of a scalar ergodic diffusion X , given continuous observations $(X_t)_{0 \leq t \leq T}$, can be estimated under standard nonparametric assumptions with a parametric rate (see, e.g., Chapter 4.2 in [13]). One further example which is relevant from a statistical point of view concerns uniform CLTs, that is, Donsker-type theorems. Van der Vaart and van Zanten [27] have shown indeed that there is no gap between the pre-Gaussian and the Donsker property in the framework of continuous-time observations of a scalar ergodic diffusion process with finite speed measure: In contrast to the classical case of empirical processes based on i.i.d. observations, the existence of a tight Gaussian limit is not only necessary but already sufficient for a CLT which holds uniformly over an entire (possibly infinite-dimensional) class of functions. The proofs of both remarkable results rely on the existence of diffusion local time and its properties. On the one hand, this suggests that it pays off to take probabilistic properties of diffusion processes into account as there exist situations where “diffusion specific” results which go beyond the well-known results established in the i.i.d. framework can be obtained. On the other hand, due to the use of diffusion local time, the proofs and results mentioned above are restricted to the one-dimensional setting.

In the sequel, we want to contribute to the statistical investigation of diffusion processes in higher dimension, taking explicitly into account probabilistic properties in the multivariate framework. Precisely, we consider the question of nonparametric estimation of the invariant density of a diffusion \mathbf{X} , given as a (weak or strong) solution of the SDE

$$(1.1) \quad dX_t = b(X_t) dt + dW_t, \quad X_0 = \xi,$$

where $b: \mathbb{R}^d \rightarrow \mathbb{R}^d$ is the unknown drift vector, $W = (W_t)_{t \geq 0}$ is some d -dimensional Brownian motion and $\xi \in \mathbb{R}^d$ denotes a random vector independent of W . Throughout the sequel, we assume that the drift vector is of the form $b = -\nabla V$, denoting by $V \in C^2(\mathbb{R}^d)$ some potential function. Furthermore, we assume that \mathbf{X} is ergodic with invariant measure $\mu = \mu_V$ of the form $d\mu_V = \rho_V d\lambda$ and strictly stationary, that is, $\xi \sim \mu_V$.

The aim of this paper is to study the basic problem of estimating the invariant density from a continuous record of observations for a class of ergodic multidimensional diffusion processes. Let us emphasise that we consider this specific question as a practical test for a multidimensional diffusion model. Indeed, we will

see that the technical devices which are applicable in the identified model allow to expose “diffusion specific” statistical properties (in the given case: faster rates of convergence) also in higher dimension. An important tool in our investigation are functional inequalities which allow to incorporate specific probabilistic findings on diffusion processes. Precisely, the model (1.1) will be studied under Assumptions $(\mathcal{M}1)$ – $(\mathcal{M}4)$ to be introduced below.

Adaptive estimation with respect to sup-norm loss under anisotropic Hölder smoothness constraints. We consider adaptive estimation of the invariant density $\rho = \rho_V$ of the diffusion \mathbf{X} , quantifying the quality of the estimation procedure $\widehat{\rho}_T$ by the sup-norm risk,

$$(1.2) \quad \begin{aligned} \mathcal{R}_T^{(q)}(\widehat{\rho}_T, \rho) &:= \left(\mathbf{E}_b^{(T)} \left[\left(\sup_{x \in \mathbb{R}^d} |\widehat{\rho}_T(x) - \rho(x)| \right)^q \right] \right)^{1/q} \\ &= \left(\mathbf{E}_b^{(T)} [\|\widehat{\rho}_T - \rho\|_\infty^q] \right)^{1/q}, \quad q \geq 1, T > 0, \end{aligned}$$

where $\mathbf{E}_b^{(T)}$ is the expectation with respect to the law $\mathbf{P}_b^{(T)}$ of the sample $X^T := (X_t)_{0 \leq t \leq T}$ of the diffusion process solution of (1.1). Generally speaking, the literature on statistical properties of multidimensional diffusion processes is still relatively scarce. An important and substantial point of reference is given by the results of [7] on asymptotic statistical equivalence for inference on the drift in the multidimensional diffusion case. As a by-product of the study, [7] prove convergence rates of invariant density estimators for pointwise estimation under isotropic Hölder smoothness constraints which are considerably faster than those known from standard multivariate density estimation. This result relies on “small” upper bounds on the variance of additive diffusion functionals, proven by an application of the spectral gap inequality in combination with a bound on the transition densities of the diffusion. We construct *adaptive* estimators in the multidimensional diffusion case which achieve “fast” rates of convergence with respect to (1.2) over *anisotropic* Hölder balls.

The extension of results on pointwise rates of convergence to the sup-norm loss usually is fairly standard. It is nevertheless unclear from the outset how the improved variance estimates can be applied in an effective manner in a data-driven estimation procedure. Furthermore, the notion of anisotropy plays an important role in our investigation. Throughout, we consider situations where the smoothness properties of elements of a function space may depend on the chosen direction in \mathbb{R}^d . The Russian school considered anisotropic spaces from the very start of the theory of function spaces in the late 1950s and the early 1960s (see, e.g., [23] for an account of the developments). Nonetheless, results on minimax rates of convergence in classical statistical models were rare for a long time. The question of optimal bandwidth selection for density estimation (based on i.i.d. observations) with respect to sup-norm risk was not completely solved until the rather recent developments in [15]. Lepski’s (convolution-operator based) selection procedure actually provides the starting point of our study.

The general diffusion framework. We consider the following diffusion model.

DEFINITION 1.1. Given $\mathcal{C}, \mathcal{C}_0, \mathcal{C}_1, \mathcal{C}_2 > 0$, denote by $\Sigma = \Sigma(\mathcal{C}, \mathcal{C}_0, \mathcal{C}_1, \mathcal{C}_2)$ the set of all functions $b = -\nabla V : \mathbb{R}^d \rightarrow \mathbb{R}^d, V \in C^2(\mathbb{R}^d)$, such that:

(M1) (1.1) admits a unique weak solution \mathbf{X} with associated Markov semigroup $(P_t)_{t \geq 0}$ and infinitesimal generator L ;

(M2) \mathbf{X} is ergodic with unique invariant probability measure μ_V which is absolutely continuous with respect to the Lebesgue measure λ and satisfies $d\mu_V = \rho_V d\lambda$, with uniformly bounded invariant density $\|\rho_V\|_\infty \leq \mathcal{C}$ which is given by

$$(1.3) \quad \rho_V(x) := \left(\int_{\mathbb{R}^d} e^{-2V(u)} du \right)^{-1} \exp(-2V(x)), \quad x \in \mathbb{R}^d;$$

(M3) the diffusion \mathbf{X} satisfies the Poincaré inequality [see (PI) in Definition 2.1 below] with constant \mathcal{C}_0 ; and

(M4) the diffusion \mathbf{X} satisfies the Nash inequality [see (NI) in Definition 2.1 below] with constants \mathcal{C}_1 and \mathcal{C}_2 .

Denote by $\mathbb{H}_d(\boldsymbol{\beta}, \mathcal{L})$ the set of invariant density functions $\rho = \rho_V$ which are associated with drift coefficients $b \in \Sigma$ and which belong to *anisotropic Hölder balls* (see Definition 3.1) of regularity $\boldsymbol{\beta} + \mathbf{1} = (\beta_1 + 1, \dots, \beta_d + 1)$ and radius $\mathcal{L} = (\mathcal{L}_1, \dots, \mathcal{L}_d)$ ($\beta_r, \mathcal{L}_r > 0, r = 1, \dots, d$).

Let us already remark that a classical result (cf. [22]) states that the assumed gradient structure of the drift (i.e., the existence of a smooth potential function $V : \mathbb{R}^d \rightarrow \mathbb{R}$) is equivalent to self-adjointness and thus reversibility of the generator L in $L^2(\mu_V)$. Reversibility again permits to use nice analytical results for the associated Markov semigroup. Conditions (M3) and (M4) have been included in Definition 1.1 for two reasons: On the one hand, Holley–Stroock classical perturbation arguments provide a stable framework which is required for the proof of *lower bounds* on the rate of convergence of estimators. With regard to the proof of *upper bounds* on the convergence rate, (M3) and (M4) allow to establish tight upper bounds on the variance

$$\text{Var}_b \left(\int_0^T f(X_u) du \right), \quad f \in L^2(\mu_V),$$

of integral functionals of the diffusion \mathbf{X} . Bounds of this type were proven before in [7] (cf. their Proposition 1) by combining estimates based on the spectral gap inequality and on upper bounds on the transition densities of \mathbf{X} . As already pointed out, in view of the statistical minimax setup considered in this paper, it appears convenient to work in the (stable) framework of functional inequalities. The Nash inequality (NI) actually allows to derive convenient upper bounds on the transition operator. Given any bounded function $f : \mathbb{R}^d \rightarrow \mathbb{R}$ with support \mathcal{S} satisfying

$\lambda(\mathcal{S}) < 1$, one might then specify an explicit constant $C = C(d, \mathcal{C}_0, \mathcal{C}_1, \mathcal{C}_2)$ such that, for any $T > 0$,

$$(1.4) \quad \text{Var}_b \left(\int_0^T f(X_u) \, du \right) \leq CT \|f\|_{L^1(\mu_V)} \|f\|_\infty \lambda(\mathcal{S}) \psi_d^2(\lambda(\mathcal{S})),$$

where

$$(1.5) \quad \psi_d(x) := \begin{cases} 1, & d = 1, \\ 1 \vee (\log(1/x))^{1/2}, & d = 2, \\ x^{(1/d)-(1/2)}, & d \geq 3, \end{cases} \quad x > 0.$$

For a more precise statement of the result and its proof, we refer to Section B in the Appendix (see Lemma B.1).

The collection of invariant density estimators. Given the observation X^T of a diffusion \mathbf{X} as described in Definition 1.1, we propose to estimate the invariant density ρ_V by means of a kernel estimator. For doing so, consider some symmetric, Lipschitz-continuous kernel function $K : \mathbb{R} \rightarrow \mathbb{R}$ with support $\text{supp}(K)$ satisfying

$$(K) \quad \int K \, d\lambda = 1, \quad k_\infty := \|K\|_\infty < \infty \quad \text{and} \quad \text{supp}(K) \subset [-1/2, 1/2].$$

Denoting by $X_t^j, j \in \{1, \dots, d\}$, the j th component of $X_t, t \geq 0$, a natural estimator of ρ_V at $x = (x_1, \dots, x_d)^\top \in \mathbb{R}^d$ in the anisotropic context is given by

$$(1.6) \quad \hat{\rho}_{T, \mathbf{h}}(x) := \frac{1}{T \prod_{l=1}^d h_l} \int_0^T \prod_{m=1}^d K \left(\frac{x_m - X_u^m}{h_m} \right) \, du,$$

and the main question concerns the choice of the multi-index bandwidth $\mathbf{h} = (h_1, \dots, h_d)^\top$.

Let us first comment on the special case of scalar and two-dimensional diffusion processes. Using any sufficiently regular kernel function K , it is straightforward to show that the “universal” bandwidth choice $h_1 = h_2 \sim T^{-1/2}$ yields for any $q \geq 1$ the rates

$$(1.7) \quad \sup_{\rho \in \mathbb{H}_d(\beta, \mathcal{L})} \mathcal{R}_T^{(q)}(\hat{\rho}_T, \rho) = \begin{cases} O(\sqrt{\log T/T}), & d = 1, \\ O(\log T/\sqrt{T}), & d = 2 \end{cases}$$

(cf. Section 3 for details). Both the bandwidth and the upper bound on the rate of convergence appearing on the right hand side of (1.7) do *not* depend on the unknown smoothness of the invariant density ρ_V . Thus, there is no need for proposing a data-driven bandwidth selection procedure in the framework of continuous observations of a lower-dimensional diffusion process. In what follows, we therefore restrict to the question of adaptive invariant density estimation in dimension $d \geq 3$.

The basic approach is briefly summarised as follows: Specify the set \mathcal{H}_T of candidate bandwidths as

$$(1.8) \quad \mathcal{H}_T := \left\{ \mathbf{h} = (h_1, \dots, h_d)^\top \in (0, 1]^d : T \alpha_\circ^2 \geq \prod_{j=1}^d h_j^{(2/d)-1} \log T \right\},$$

where $\alpha_\circ := (2\gamma_{2q}^\circ(d, k_\infty; \mathcal{C}_0, \mathcal{C}_1, \mathcal{C}_2))^{-2}$, for $q \geq 1$ used in the definition of the risk criterion [see (1.2)] and some function γ_p° taking values in \mathbb{R}_+ whose precise description is postponed to Remark 3.5. Introduce the family of estimators

$$(1.9) \quad \mathcal{F}(\mathcal{H}_T) := \left\{ \widehat{\rho}_{\mathbf{h}}(x) := \frac{1}{T} \int_0^T \mathbb{K}_{\mathbf{h}}(X_u - x) du : x \in \mathbb{R}^d, \mathbf{h} \in \mathcal{H}_T \right\},$$

where

$$(1.10) \quad \mathbb{K}_{\mathbf{h}}(u) := V_{\mathbf{h}}^{-1} \prod_{m=1}^d K(u_m/h_m) \quad \text{for } V_{\mathbf{h}} := \prod_{l=1}^d h_l.$$

The aim now is to select an estimator from the family $\mathcal{F}(\mathcal{H}_T)$ in a completely data-driven way, based only on the observation of the continuous trajectory X^T .

Statement of main results. We propose a procedure to select an estimator from $\mathcal{F}(\mathcal{H}_T)$ which is motivated by the approach of [15]. In particular, it will be proven that the selected estimator $\widehat{\rho}_{\widehat{\mathbf{h}}} \in \mathcal{F}(\mathcal{H}_T)$ satisfies, for any $T > 0$ and any $b \in \Sigma$,

$$(1.11) \quad \mathcal{R}_T^{(q)}(\widehat{\rho}_{\widehat{\mathbf{h}}}, \rho_V) \leq C_1 \inf_{\mathbf{h} \in \overline{\mathcal{H}}_T} \left\{ \mathcal{B}_V(\mathbf{h}) + \prod_{j=1}^d h_j^{(1/d)-(1/2)} \sqrt{\frac{\log T}{T}} \right\} + \frac{C_2}{\sqrt{T}}.$$

Here, C_1 and C_2 are numerical constants depending on $\mathcal{F}(\mathcal{H}_T)$ and the set Σ only, $\overline{\mathcal{H}}_T \subset \mathcal{H}_T$ denotes a dyadic grid, and $\mathcal{B}_V(\cdot)$ can be viewed as the approximation error of ρ_V measured in the supremum norm. Inequality (1.11) then serves as the main tool for analysing the rate of convergence of the adaptive estimators $\widehat{\rho}_{\widehat{\mathbf{h}}}$.

Let $\boldsymbol{\beta} = (\beta_1, \dots, \beta_d)$, $\beta_r > 0$, and define the *mean smoothness* $\overline{\boldsymbol{\beta} + \mathbf{1}}$ of $\boldsymbol{\beta} + \mathbf{1}$ by letting

$$\overline{\boldsymbol{\beta} + \mathbf{1}} = \frac{1}{d} \sum_{r=1}^d \frac{1}{\beta_r + 1}.$$

Given any $d \geq 3$, introduce further

$$(1.12) \quad \begin{aligned} \Psi(\boldsymbol{\beta} + \mathbf{1}) &= \Psi(\beta_1 + 1, \dots, \beta_d + 1) := \frac{\overline{\boldsymbol{\beta} + \mathbf{1}}}{2\overline{\boldsymbol{\beta} + \mathbf{1}} + d - 2}, \\ \varphi_T(\boldsymbol{\beta} + \mathbf{1}) &= \varphi_T(\beta_1 + 1, \dots, \beta_d + 1) := \left(\frac{\log T}{T} \right)^{\Psi(\boldsymbol{\beta} + \mathbf{1})}. \end{aligned}$$

We show that, for any $q \geq 1$,

$$(1.13) \quad \limsup_{T \rightarrow \infty} \sup_{\rho \in \mathbb{H}_d(\beta, \mathcal{L})} (\mathbf{E}_b^{(T)}[\varphi_T^{-q}(\beta + \mathbf{1}) \|\widehat{\rho}_{\mathbf{h}} - \rho\|_{\infty}^q])^{1/q} < \infty.$$

(See Theorems 3.3 and 3.4 in Section 3.1 for the precise statement of the results.) We further establish a lower bound, proving optimal adaptivity of the suggested estimators over a scale of anisotropic Hölder classes. The optimal rate of convergence for density estimation with respect to sup-norm risk over anisotropic Hölder classes, given n \mathbb{R}^d -valued i.i.d. observations of smoothness $(\beta_1 + 1, \dots, \beta_d + 1)$, $\beta_r > 0$, is given as

$$\left(\frac{\log n}{n}\right)^{\frac{1}{2 + \sum_{j=1}^d \frac{1}{\beta_j + 1}}} = \left(\frac{\log n}{n}\right)^{\frac{\overline{\beta+1}}{2\overline{\beta+1} + d}}$$

(see Theorems 2 and 3 in [15]). The rate $(\log T/T)^{\overline{\beta+1}/(2\overline{\beta+1} + d - 2)}$ in (1.13) reflects the increased smoothness in the framework of continuous observations and the specific properties of Markovian diffusion semigroups. Depending on the degree of anisotropy, it may turn out to be substantially faster than the isotropic special case $(\log T/T)^{(\beta+1)/(2\beta+d)}$.

Outline of the paper. Preliminaries and central auxiliary results are introduced in Section 2. In particular, we formulate implications of the functional inequalities (of Poincaré- and Nash-type, resp.) which we take for granted throughout the sequel. A reader who is mostly interested in the concrete statistical application might skip this section and continue directly with reading Section 3. There, we investigate the question of estimating the invariant density from continuous observations under anisotropic Hölder smoothness constraints. The main results are given in Section 3.3: We formulate the sup-norm oracle-type inequality (1.11) which is at the heart of the adaptive estimation scheme. In addition, we state an asymptotic lower bound which confirms optimality of the adaptive results. In Section 4, we study issues related to invariant density estimation based on discrete observations. On the one hand, we consider convergence properties of estimators in the high-frequency framework, and we specify at which rate the sampling frequency has to tend to zero in order to retain the “fast” convergence rate which has been found in the setup of continuous observations. On the other hand, we sketch a concrete algorithm which yields data-driven invariant density estimators based on discrete observations. In addition, we state the low-frequency counterpart of (1.13). Complements on functional inequalities and the derivation of upper functions for positive random functionals in the diffusion setup are given in the Appendix. All remaining proofs are deferred to the Supplementary Material [26].

2. Preliminaries. Let $\mathbf{X} = (X_t)_{t \geq 0}$ be a diffusion on a probability space $(\Omega, \mathfrak{A}, \mathbf{P})$, taking values in \mathbb{R}^d . Define the associated transition semigroup $(P_t)_{t \geq 0}$ on suitable measurable functions $f: \mathbb{R}^d \rightarrow \mathbb{R}$ by setting

$$P_t f(x) = \mathbf{E}[f(X_t) \mid X_0 = x], \quad t \geq 0, x \in \mathbb{R}^d.$$

We always assume that \mathbf{X} is ergodic with invariant probability measure μ such that $(P_t)_{t \geq 0}$ defines a semigroup of contractions on $L^2(\mu)$. Throughout, $L^p(\mu) = L^p(\mathbb{R}^d, \mu)$, $1 \leq p \leq \infty$, denote the Lebesgue spaces on the measure space $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \mu)$ with associated norm $\|\cdot\|_{L^p(\mu)} = \|\cdot\|_p$.

It follows from the Hille–Yosida theory that there is a dense linear subspace of the Hilbert space $L^2(\mu)$, called the *domain* \mathbb{D} of the semigroup $(P_t)_{t \geq 0}$, on which the derivative at $t = 0$ of P_t exists in $L^2(\mu)$. The operator that maps $f \in \mathbb{D}$ to this derivative $L f \in L^2(\mu)$ of $P_t f$ at $t = 0$ is a linear operator, the so-called *infinitesimal generator* of the semigroup $(P_t)_{t \geq 0}$. The generator L and its domain $\mathbb{D}(L)$ completely characterise the semigroup $(P_t)_{t \geq 0}$ acting on $L^2(\mu)$. Throughout the sequel, we restrict to symmetric Markov diffusion semigroups and the associated reversible diffusion processes. Recall that the Markov semigroup $(P_t)_{t \geq 0}$ is said to be *symmetric* with respect to the invariant measure μ [or: μ is *reversible* for $(P_t)_{t \geq 0}$] if, for all $f, g \in L^2(\mu)$ and all $t \geq 0$, $\int f P_t g \, d\mu = \int g P_t f \, d\mu$.

In order to clarify the statistical investigation, we further restrict to diffusion processes with unit diffusion matrix. In this case, it is a classical result (cf. [22]) that the assumed reversibility of the diffusion \mathbf{X} in $L^2(\mu)$ is equivalent to a gradient structure of the drift, that is, there exists a smooth potential function $V: \mathbb{R}^d \rightarrow \mathbb{R}$ such that

$$L f(x) = -\langle \nabla V(x), \nabla f(x) \rangle + \frac{1}{2} \Delta f(x), \quad x \in \mathbb{R}^d, f \in \mathbb{D}(L).$$

Assume that we are given a vector subspace \mathcal{A} of the domain $\mathbb{D}(L)$ such that, for every pair $(f, g) \in \mathcal{A} \times \mathcal{A}$, the product $f g$ is in the domain $\mathbb{D}(L)$. The bilinear map

$$\Gamma(f, g) := \frac{1}{2} [L(fg) - f Lg - g Lf], \quad (f, g) \in \mathcal{A} \times \mathcal{A},$$

is called the *carré du champ operator* of L . In the given setup, it takes the form $\Gamma(f, g) = \langle \nabla f, \nabla g \rangle$ for smooth functions f, g on \mathbb{R}^d . In general, the carré du champ operator Γ and the measure μ completely determine the symmetric Markov generator L .

2.1. Functional inequalities. There exists a very large family of functional inequalities for symmetric Markov semigroups which is subject of many recent works. For a comprehensive account on related issues in the specific context of Markov diffusion operators, we refer to [3]. Many functional inequalities—such as the Poincaré (or spectral gap) inequality as the most prominent example—compare L^p norms of functions to the L^2 norms of their gradients. (As noted in [3], the terminology between Poincaré and spectral gap inequality oscillates.) An important

feature of Poincaré inequalities is their equivalence to exponential convergence of the associated Markov semigroup in $L^2(\mu)$. We shall further use Nash inequalities which belong to the broad class of Sobolev inequalities. Their main interest is that they easily provide good control on heat kernels. For ease of presentation, we include the following formal definition.

DEFINITION 2.1. The generator L (or its carré du champ Γ) is said to satisfy a *Poincaré inequality* on \mathcal{A} with respect to the probability measure μ with constant $C_0 > 0$ if, for any $f \in \mathcal{A}$,

$$(PI) \quad \text{Var}_\mu(f) := \int f^2 d\mu - \left(\int f d\mu \right)^2 \leq C_0 \int \Gamma(f, f) d\mu.$$

The generator L (or its carré du champ Γ) is said to satisfy a *Nash inequality* on \mathcal{A} with respect to the probability measure μ with constants $C_1 \geq 0, C_2 > 0$ if, for any $f \in \mathcal{A}$,

$$(NI) \quad \|f\|_{L^2(\mu)}^{d+2} \leq \left(C_1 \|f\|_{L^2(\mu)}^2 + C_2 \int \Gamma(f, f) d\mu \right)^{d/2} \|f\|_{L^1(\mu)}^2.$$

In the symmetric framework, (PI) is known to be equivalent to exponential convergence of the semigroup $(P_t)_{t \geq 0}$ to equilibrium in $L^2(\mu)$: (PI) holds if and only if for every function $f: \mathbb{R}^d \rightarrow \mathbb{R}$ in $L^2(\mu)$,

$$(2.1) \quad \text{Var}_\mu(P_t f) \leq e^{-2t/C_0} \text{Var}_\mu(f)$$

(see Theorem 4.2.5, page 183, in [3]).

REMARK 2.2. For later reference, let us also recall some classical findings concerning the family of Nash-type inequalities. Whenever μ is a probability measure, it holds $C_1 \geq 1$ in (NI). By way of contrast, one may have $C_1 = 0$ in \mathbb{R}^d with the Lebesgue measure λ . In particular, this is the case for the *classical Nash inequality* (cf. [21]) which may be stated as

$$(NI_0) \quad \|f\|_{L^2(\lambda)}^{1+(d/2)} \leq C_d \|f\|_{L^1(\lambda)} \|\nabla f\|_{L^2(\lambda)}^{d/2},$$

for any sufficiently smooth $f: \mathbb{R}^d \rightarrow \mathbb{R}$. Whenever μ is a probability measure and $C_1 = 1$, (NI) implies a Poincaré inequality (PI). (See Proposition 6.2.2 and page 282 in [3].) A Nash inequality (NI) holds if and only if the operator P_t is bounded from $L^1(\mu)$ into L^∞ with norm bounded above by $a + bt^{-d/2}$. For the classical Euclidean Nash inequality (NI₀), the equivalence is valid with a bound of the form $at^{-d/2}$ (cf. Theorem 2.2 in [1]).

2.2. *The statistical diffusion model.* Recall that our principal object of statistical investigation are diffusion processes \mathbf{X} , described by the SDE

$$(2.2) \quad dX_t = b(X_t) dt + dW_t, \quad 0 \leq t \leq T,$$

with (unknown) drift vector $b = -\nabla V: \mathbb{R}^d \rightarrow \mathbb{R}^d$ and some d -dimensional Wiener process $W = (W_t)_{t \geq 0}$. Assuming that \mathbf{X} is ergodic with unique invariant probability measure μ_V which is absolutely continuous with respect to the Lebesgue measure ($d\mu_V = \rho_V d\lambda$), we will consider the question of estimating the invariant density ρ_V under (nonparametric) smoothness constraints. In applications, one of course usually is confronted with the question of finding easily verifiable criteria which ensure the existence of a solution of the SDE (2.2), ergodicity or the validity of certain functional inequalities. A brief account on this last issue is given in Appendix A.

For the moment, we content ourselves with the following example.

EXAMPLE 2.3 (cf. Section 7.7 in [3]). Consider the family of probability measures

$$d\mu_\alpha(x) = c_\alpha \exp(-(1 + \|x\|^2)^{\alpha/2}) dx, \quad x \in \mathbb{R}^d, \alpha > 0,$$

$c_\alpha > 0$ the normalising constant. For any $\alpha > 0$, the associated Markov semigroup $(P_t)_{t \geq 0}$ has density kernels $p_t(x, y)$, $t > 0$, $(x, y) \in \mathbb{R}^d \times \mathbb{R}^d$, with respect to μ_α , satisfying

$$P_t f(x) = \int p_t(x, y) f(y) dy,$$

$$p_t(x, y) \leq C(t)w(x)w(y) \quad \text{with } w(\cdot) := \exp(-(1 + \|\cdot\|^2)^{\alpha/2}/2)$$

and $C(t) > 0$. For any $\alpha \geq 1$, a Poincaré inequality (PI) with constant $C_0 > 0$ holds. For $\alpha > 2$, the semigroup $(P_t)_{t \geq 0}$ is ultracontractive, that is, P_t is bounded from $L^1(\mu)$ to L^∞ for any $t > 0$, and the density kernels p_t are uniformly bounded for every $t > 0$. In particular, a Nash inequality (NI) holds (see Theorem 6.3.1 and Proposition 6.2.3 in [3] for the proof of the corresponding implications).

REMARK 2.4. The diffusion model introduced in Definition 1.1 is very close to the diffusion experiment considered in [7]. In this article, asymptotic local equivalence for inference on the drift of multidimensional ergodic diffusions is established for a class of diffusion processes solving (2.2) with drift coefficient $b = -\nabla V: \mathbb{R}^d \rightarrow \mathbb{R}^d$ under the following assumption: There exist constants $M_1, M_2 > 0$ such that, for any $x, y \in \mathbb{R}^d$:

- (I) $\|b(x)\| \leq M_1(1 + \|x\|)$, and
- (II) $\langle b(x) - b(y), x - y \rangle \leq -M_2\|x - y\|^2$.

Condition **(II)** in particular implies that (2.2) has a unique strong solution \mathbf{X} . In addition, \mathbf{X} is ergodic with unique invariant probability measure, and **(PI)** holds. The assumption **(I)** of at most linear growth further entails that there exists a constant $C_0 > 0$ such that the transition density $p_t(x, y)$ for all $t > 0$ and for all $(x, y) \in \mathbb{R}^d \times \mathbb{R}^d$ with $\|x - y\|^2 < t$ is bounded as follows:

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

Replacing $\psi_d(\cdot)$ defined in (1.5) with

$$\psi_d^\circ(x) := \begin{cases} \psi_1(x), & d = 1, \\ 1 \vee (\log(1/x))^2, & d = 2, \\ \psi_d(x), & d \geq 3, \end{cases} \quad x > 0,$$

a similar variance estimate as in (1.4) holds for any bounded $f: \mathbb{R}^d \rightarrow \mathbb{R}$ with support \mathcal{S} such that $\lambda(\mathcal{S}) < 1$.

3. Adaptive estimation of the invariant density in higher dimensions.

Suppose we are given a continuous record of observations $X^T \equiv (X_t)_{0 \leq t \leq T}$, $T > 0$, of a multidimensional diffusion \mathbf{X} as described in Definition 1.1. The question of density estimation belongs to the canonical framework of nonparametric statistics. However, even in the classical i.i.d. setup, the (more involved) definition of adaptive density estimators is substantially less studied than the Gaussian white noise or the nonparametric regression model. As mentioned in the Introduction and detailed below, data-driven bandwidth selection procedures are not required for estimating the invariant density of one- or two-dimensional diffusions from continuous observations. We thus focus on the question of adaptive invariant density estimation in dimension $d \geq 3$ for which, to the best of our knowledge, no results are available.

In most cases, the regularity of some function $g: \mathbb{R}^d \rightarrow \mathbb{R}$ depends on the chosen direction in \mathbb{R}^d , and isotropic spaces actually are anisotropic spaces. Depending on the specification, the consideration of anisotropy might result in a substantial dimension reduction. We thus work under the following anisotropic smoothness constraints.

DEFINITION 3.1. Let $\beta = (\beta_1, \dots, \beta_d)$, $\beta_i > 0$, $\mathcal{L} = (\mathcal{L}_1, \dots, \mathcal{L}_d)$, $\mathcal{L}_i > 0$. A function $g: \mathbb{R}^d \rightarrow \mathbb{R}$ is said to belong to the *anisotropic Hölder class* $\mathcal{H}_d(\beta, \mathcal{L})$ of functions if, for all $i = 1, \dots, d$,

$$\begin{aligned} \|D_i^k g\|_\infty &\leq \mathcal{L}_i \quad \forall k = 0, 1, \dots, \lfloor \beta_i \rfloor, \\ \|D_i^{\lfloor \beta_i \rfloor} g(\cdot + t\mathbf{e}_i) - D_i^{\lfloor \beta_i \rfloor} g(\cdot)\|_\infty &\leq \mathcal{L}_i |t|^{\beta_i - \lfloor \beta_i \rfloor} \quad \forall t \in \mathbb{R}, \end{aligned}$$

for $D_i^k g$ denoting the k th order partial derivative of g with respect to the i th component, $\lfloor \beta \rfloor$ denoting the largest integer strictly smaller than β and $\mathbf{e}_1, \dots, \mathbf{e}_d$ denoting the canonical basis in \mathbb{R}^d . Given $C, C_0, C_1, C_2 > 0$, denote by $\mathbb{H}_d(\beta, \mathcal{L}) =$

$\mathbb{H}_d(\boldsymbol{\beta}, \mathcal{L}; \mathcal{C}, \mathcal{C}_0, \mathcal{C}_1, \mathcal{C}_2)$ the set of functions $\rho_V \in \mathcal{H}_d(\boldsymbol{\beta} + \mathbf{1}, \mathcal{L})$ with $b = -\nabla V \in \Sigma(\mathcal{C}, \mathcal{C}_0, \mathcal{C}_1, \mathcal{C}_2)$.

Pointwise estimation revisited. We start our investigation of convergence rates for invariant density estimation by revisiting the remarkable result of [7] on upper bounds for *pointwise* rates and extending it to the anisotropic situation. Given positive constants M_1, M_2 , denote by $\Sigma(M_1, M_2)$ the class of drift coefficients $b = -\nabla V: \mathbb{R}^d \rightarrow \mathbb{R}^d$ satisfying conditions **(I)** and **(II)** in Remark 2.4. For $\boldsymbol{\beta}, \mathcal{L} \in (0, \infty)^d$, denote by $\tilde{\mathbb{H}}_d(\boldsymbol{\beta}, \mathcal{L}; M_1, M_2) = \tilde{\mathbb{H}}_d(\boldsymbol{\beta}, \mathcal{L})$ the set of functions $\rho_V \in \mathcal{H}_d(\boldsymbol{\beta} + \mathbf{1}, \mathcal{L})$ with $b = -\nabla V \in \Sigma(M_1, M_2)$.

In order to estimate some invariant density $\rho_V \in \tilde{\mathbb{H}}_d(\boldsymbol{\beta}, \mathcal{L})$ at the point $x = (x_1, \dots, x_d)^\top \in \mathbb{R}^d$, recall the definition of the kernel estimator $\widehat{\rho}_{T, \mathbf{h}}(\cdot)$ according to (1.6), $\mathbf{h} = (h_1, \dots, h_d)^\top$ denoting the multi-bandwidth. Suppose that the involved kernel $K: \mathbb{R} \rightarrow \mathbb{R}$ is sufficiently smooth. Assuming *isotropic* Hölder-smoothness of ρ_V of order $\beta + 1$, it is shown in Corollary 1 in [7] that the bandwidth choice $h_i \sim T^{-1/(2\beta+d)}, i = 1, \dots, d$, implies that

$$(3.1) \quad \mathbf{E}_b^{(T)} [(\widehat{\rho}_{T, \mathbf{h}}(x) - \rho_V(x))^2] \lesssim \begin{cases} T^{-1}(\log T)^4, & d = 2, \\ T^{-\frac{2(\beta+1)}{2\beta+d}}, & d \geq 3. \end{cases}$$

Extending (3.1) to the case of anisotropic Hölder smoothness, we obtain for $\rho_V \in \tilde{\mathbb{H}}_d(\boldsymbol{\beta}, \mathcal{L})$ the risk estimates

$$\mathbf{E}_b^{(T)} [(\widehat{\rho}_{T, \mathbf{h}}(x) - \rho_V(x))^2] \lesssim \sum_{\ell=1}^d h_\ell^{2(\beta_\ell+1)} + T^{-1}(\psi_d^\circ)^2 \left(\prod_{\ell=1}^d h_\ell \right).$$

For a suitably chosen multi-bandwidth \mathbf{h} and kernel K , this yields the following upper bound on the mean-squared error:

$$(3.2) \quad \mathbf{E}_b^{(T)} [(\widehat{\rho}_{T, \mathbf{h}}(x) - \rho_V(x))^2] \lesssim \begin{cases} T^{-1}(\log T)^4, & d = 2, \\ T^{-\frac{2\beta+1}{2\beta+1+d-2}}, & d \geq 3. \end{cases}$$

Under the assumptions imposed in Definition 1.1, the upper bound in (3.2) for $d = 2$ can be improved to $T^{-1} \log T$.

Invariant density estimation in the one- and two-dimensional case. Let $K: \mathbb{R} \rightarrow \mathbb{R}$ be a symmetric, Lipschitz-continuous kernel satisfying **(K)**, and, for some integer $\mathfrak{b} \geq 2$,

$$(3.3) \quad \int_{\mathbb{R}} u^m K(u) du = 0 \quad \forall m = 1, \dots, \mathfrak{b} + 1.$$

If \mathbf{X} is strictly stationary, it can be shown by standard arguments that the kernel estimator $\widehat{\rho}_{T, \mathbf{h}}(\cdot)$ defined in (1.6) satisfies, for any $\boldsymbol{\beta} \in (0, \mathfrak{b}]^d, \mathcal{L} \in (0, \infty)^d, d \in$

{1, 2} and $\psi_d(\cdot)$ introduced in (1.5),

$$\sup_{\rho \in \mathbb{H}_d(\beta, \mathcal{L})} \mathbf{E}_b^{(T)}[\|\widehat{\rho}_{T, \mathbf{h}} - \rho\|_\infty] = O\left(\sum_{j=1}^d h_j^{\beta_j+1} + \psi_d\left(\prod_{j=1}^d h_j\right)\sqrt{\frac{\log T}{T}}\right).$$

Inserting $h_1 = h_2 \sim T^{-1/2}$ into the right hand side of the last display, one obtains the rate announced in (1.7) in the **Introduction**. There is obviously no gain in implementing a data-driven bandwidth selection procedure for density estimation in the framework of continuous observations of a one- or two-dimensional diffusion process. Throughout the sequel, we therefore restrict to the case $d \geq 3$. With regard to the practically most relevant case of adaptive invariant density estimation from low-frequency observations in the scalar diffusion case, we refer to the recent work of [25] who, in addition, construct adaptive confidence bands for the drift coefficient by using a data-driven estimator.

3.1. Selection rule. Consider some symmetric, Lipschitz-continuous kernel function $K : \mathbb{R} \rightarrow \mathbb{R}$ satisfying **(K)**. Throughout the sequel, we denote $k_1 := \|K\|_{L^1(\lambda)}$. Recall the definition of the set of candidate bandwidths \mathcal{H}_T and the associated set $\mathcal{F}(\mathcal{H}_T)$ of kernel estimators in (1.8) and (1.9), respectively. We now turn to describing a selection procedure from $\mathcal{F}(\mathcal{H}_T)$. Later on, we show that this rule yields minimax adaptive invariant density estimators over a scale of anisotropic Hölder classes.

Auxiliary quantities. Following [15], the selection rule is based on auxiliary estimators relying on the convolution operator. To the best of our knowledge, the use of convolution-based auxiliary estimators was pioneered in [19]. It was introduced as a device to circumvent the lack of ordering among a set of estimators in anisotropic situations where the increase of the variance of an estimator does not necessarily imply a decrease of its bias.

For any bandwidths $\mathbf{h} = (h_1, \dots, h_d)^\top, \boldsymbol{\eta} = (\eta_1, \dots, \eta_d)^\top \in \mathcal{H}_T$ and any point $z = (z_1, \dots, z_d)^\top \in \mathbb{R}^d$, denote

$$(3.4) \quad (\mathbb{K}_{\mathbf{h}} \star \mathbb{K}_{\boldsymbol{\eta}})(z) := \prod_{j=1}^d (K_{h_j} * K_{\eta_j})(z_j) = \prod_{j=1}^d \int_{\mathbb{R}} K_{h_j}(u - z_j) K_{\eta_j}(u) du,$$

and define the kernel estimators

$$\widehat{\rho}_{\mathbf{h}, \boldsymbol{\eta}}(x) := \frac{1}{T} \int_0^T (\mathbb{K}_{\mathbf{h}} \star \mathbb{K}_{\boldsymbol{\eta}})(X_u - x) du, \quad x \in \mathbb{R}^d.$$

The basic idea now is to exploit the fact that, for a pair of dilated kernels $\mathbb{K}_{\mathbf{h}}, \mathbb{K}_{\boldsymbol{\eta}}$, the smallest majorant with respect to some specific partial order relation is given by $\mathbb{K}_{\mathbf{h}} + \mathbb{K}_{\boldsymbol{\eta}} - \mathbb{K}_{\mathbf{h}} \star \mathbb{K}_{\boldsymbol{\eta}}$. Taking up this idea, the proposed selection procedure relies on comparing the differences $\widehat{\rho}_{\mathbf{h}, \boldsymbol{\eta}} - \widehat{\rho}_{\boldsymbol{\eta}}$. In particular, the method uses the

fact that, for any pair $(\mathbf{h}, \boldsymbol{\eta})$ of multi-bandwidths, the sup-norm of the associated approximation error fulfills

$$(3.5) \quad \left\| \int_{\mathbb{R}^d} [\mathbb{K}_{\mathbf{h}} \star \mathbb{K}_{\boldsymbol{\eta}}](t - \cdot) \rho_V(t) dt - \int_{\mathbb{R}^d} \mathbb{K}_{\mathbf{h}}(t - \cdot) \rho_V(t) dt \right\|_{\infty} \\ \leq \mathcal{B}_V(\boldsymbol{\eta}) \int_{\mathbb{R}^d} |\mathbb{K}_{\mathbf{h}}(y)| dy = \|\mathbb{K}\|_{L^1(\lambda)} \mathcal{B}_V(\boldsymbol{\eta}),$$

where

$$\mathcal{B}_V(\boldsymbol{\eta}) := \sup_{x \in \mathbb{R}^d} \left| \int_{\mathbb{R}^d} \mathbb{K}_{\boldsymbol{\eta}}(t - x) (\rho_V(t) - \rho_V(x)) dt \right|,$$

that is, the upper bound does not depend on \mathbf{h} .

The selection rule. Denote

$$\bar{q}_T := 2 \left(1 \vee \sup_{\mathbf{h} \in \mathcal{H}_T} \left\| T^{-1} \int_0^T |\mathbb{K}_{\mathbf{h}}(X_u - \cdot)| du \right\|_{\infty} \right),$$

and introduce

$$\hat{A}_T(\mathbf{h}) := \prod_{j=1}^d h_j^{(1/d)-(1/2)} \sqrt{\frac{\bar{q}_T \log T}{T}}.$$

Denote by $\bar{\mathcal{H}}_T \subset \mathcal{H}_T$ the dyadic grid in \mathcal{H}_T , and set

$$(3.6) \quad \hat{\Delta}_T(\mathbf{h}) := \sup_{\boldsymbol{\eta} \in \bar{\mathcal{H}}_T} \{ [\|\hat{\rho}_{\mathbf{h}, \boldsymbol{\eta}} - \hat{\rho}_{\boldsymbol{\eta}}\|_{\infty} - \lambda \hat{A}_T(\boldsymbol{\eta})]_+ \},$$

for $\lambda := (1 \vee k^d) \Lambda$ and $\Lambda := \gamma_{2q}^{\circ}(d, k_{\infty}; \mathcal{C}_0, \mathcal{C}_1, \mathcal{C}_2)$. Define $\hat{\mathbf{h}}$ by setting

$$(3.7) \quad \hat{\Delta}_T(\hat{\mathbf{h}}) + \lambda \hat{A}_T(\hat{\mathbf{h}}) = \inf_{\mathbf{h} \in \bar{\mathcal{H}}_T} \{ \hat{\Delta}_T(\mathbf{h}) + \lambda \hat{A}_T(\mathbf{h}) \}.$$

The adaptive invariant density estimator is then given as

$$(3.8) \quad \hat{\rho}_{\hat{\mathbf{h}}}(x) = \frac{1}{T} \int_0^T \mathbb{K}_{\hat{\mathbf{h}}}(X_u - x) du, \quad x \in \mathbb{R}^d.$$

REMARK 3.2. Although the basic ideas underlying the selection procedure are the same as in the (anisotropic) i.i.d. framework, it appears instructive to include some further explanation of the involved quantities. More recently, convolution-based methods in adaptive estimation were suggested in [10–12] or [15]. While the proposed (invariant) density estimation procedure is much more involved, it still resembles the adaptive scheme developed by [11] in the Gaussian white noise framework. In this context, the analogue of $\hat{\Delta}_T(\mathbf{h})$ defined in (3.6)

serves as a lower estimator for the norm of the bias of the estimator $\widehat{\rho}_{\mathbf{h}}$. In our present setup, letting

$$\bar{s}_T := 1 \vee \sup_{\mathbf{h} \in \mathcal{H}_T} \left\| \int_{\mathbb{R}^d} |\mathbb{K}_{\mathbf{h}}(t - \cdot)| \rho_V(t) dt \right\|_{\infty}$$

and denoting

$$(3.9) \quad A_T(\mathbf{h}) := \prod_{j=1}^d h_j^{(1/d)-(1/2)} \sqrt{\frac{\bar{s}_T \log T}{T}},$$

it holds in fact

$$\widehat{\Delta}_T(\mathbf{h}) \leq (1 \vee k_1^d)(\Lambda A_T(\mathbf{h}) + \mathcal{B}_V(\mathbf{h})) + \text{“smaller order term”}$$

[see (C.4) in the proof of Theorem 3.3].

3.2. *Analysis of the selection rule.* With regard to the probabilistic analysis of the proposed selection procedure, the crucial point is the evaluation of upper functions for the family of centered estimators. In the current case, it is mostly reduced to verifying

$$(3.10) \quad \mathbf{E}_b^{(T)} \left[\sup_{\eta \in \mathcal{H}_T} \{ [\|\xi_{\eta}\|_{\infty} - \Lambda A_T(\eta)]_+ \} \right] \lesssim T^{-1/2}.$$

Here,

$$\xi_{\eta}(\cdot) := \widehat{\rho}_{\eta}(\cdot) - \int_{\mathbb{R}^d} \mathbb{K}_{\mathbf{h}}(t - \cdot) \rho_V(t) dt$$

denotes the stochastic term of the estimator $\widehat{\rho}_{\eta}$. To prove (3.10), we will use and modify some general results from [16, 17] and [18].

For applying them, we rely on another implication of (PI), namely the following *Bernstein-type inequality* due to [20]: For any bounded and measurable $g: \mathbb{R}^d \rightarrow \mathbb{R}$ with $\int g d\mu = 0$ and $\forall t, r > 0$, it holds

$$\begin{aligned} \mathbf{P}_{\mu} \left(\frac{1}{t} \int_0^t g(X_u) du > r \right) &\leq \exp \left(- \frac{2tr^2}{\sigma^2(g) \left(\sqrt{1 + \frac{2C_0 \|g\|_{\infty} r}{\sigma^2(g)}} + 1 \right)^2} \right) \\ (BI) \quad &\leq \exp \left(- \frac{tr^2}{2(\sigma^2(g) + C_0 \|g\|_{\infty} r)} \right), \end{aligned}$$

where $\sigma^2(\cdot)$ is the asymptotic variance in the CLT, given by

$$(3.11) \quad \begin{aligned} \sigma^2(g) &:= \lim_{t \rightarrow \infty} t^{-1} \text{Var}_{\mu} \left(\int_0^t g(X_u) du \right) \\ &= \lim_{t \rightarrow \infty} \mathbf{E}_{\mu} \left[\left(\frac{1}{\sqrt{t}} \int_0^t g(X_u) du \right)^2 \right], \end{aligned}$$

and C_0 is the constant appearing in Poincaré’s inequality (PI). The exponential inequality (BI) will be crucial for our investigation of the proposed adaptive estimators of the invariant density. There exists indeed a vast number of generalisations of the classical Bernstein inequality for independent observations to the framework of Markovian or weakly dependent processes. As a concrete example from the functional inequalities setup, [5] show that (PI) implies a deviation inequality similar to (BI), replacing $\sigma^2(g)$ with $\|g\|_{L^2(\mu)}^2$. For the proof of “fast” rates of convergence in the adaptive estimation procedure, we actually require both the tight form of Lezaud’s [20] result and the variance bounds announced in (1.4).

Coming back again to the concrete selection rule and the definition of the bandwidth $\widehat{\mathbf{h}}$ according to (3.7), it is important to note that $\widehat{A}_T(\cdot)$ presents a data-driven correspondent of $A_T(\cdot)$. It is given explicitly and does not depend on the unknown invariant density. Instead, it is completely determined by the kernel $K(\cdot)$.

3.3. *Main results.* The following inequality provides the key for analysing the statistical properties of estimators of the invariant density for diffusions as described in Definition 1.1, constructed according to the above scheme.

THEOREM 3.3. *Let $K : \mathbb{R} \rightarrow \mathbb{R}$ be a symmetric, Lipschitz-continuous kernel satisfying (K). Assume that \mathbf{X} is strictly stationary. Then, for any $q \geq 1, d \geq 3$ and $b = -\nabla V \in \Sigma(\mathcal{C}, \mathcal{C}_0, \mathcal{C}_1, \mathcal{C}_2)$, there exist $C_1, C_2 \in (0, \infty)$ such that, for any $T > 0$,*

$$(3.12) \quad \mathcal{R}_T^{(q)}(\widehat{\rho}_{\widehat{\mathbf{h}}}, \rho_V) = (\mathbf{E}_b^{(T)}[\|\widehat{\rho}_{\widehat{\mathbf{h}}} - \rho_V\|_{\infty}^q])^{1/q} \leq C_1 \mathfrak{R}_T(\rho_V) + C_2 T^{-1/2},$$

where

$$\mathfrak{R}_T(\rho_V) := \inf_{\mathbf{h} \in \overline{\mathcal{H}}_T} \left\{ \mathcal{B}_V(\mathbf{h}) + \prod_{j=1}^d h_j^{(1/d)-(1/2)} \sqrt{\frac{\log T}{T}} \right\}.$$

Given $\boldsymbol{\beta} \in (0, \infty)^d$, recall the definition of $\varphi_T(\boldsymbol{\beta} + \mathbf{1})$ according to (1.12). Inequality (3.12) does not represent a “classical” oracle inequality of the form

$$\mathcal{R}_T^{(q)}(\widehat{\rho}_{\widehat{\mathbf{h}}}, \rho_V) \leq C_1 \inf_{\mathbf{h} \in \overline{\mathcal{H}}_T} \mathcal{R}_T^{(q)}(\widehat{\rho}_{\mathbf{h}}, \rho_V) + C_2 T^{-1/2}.$$

Nonetheless, it allows to derive the following result.

THEOREM 3.4. *Let $K : \mathbb{R} \rightarrow \mathbb{R}$ be a symmetric, Lipschitz-continuous kernel satisfying (K), and, for some integer $\mathfrak{b} \geq 2$, (3.3). If \mathbf{X} is strictly stationary, the adaptive estimator $\widehat{\rho}_{\widehat{\mathbf{h}}}$ defined according to (3.8) satisfies, for any $d \geq 3, \boldsymbol{\beta} \in (0, \mathfrak{b}]^d$ and any $\mathcal{L} \in (0, \infty)^d$,*

$$(3.13) \quad \limsup_{T \rightarrow \infty} \sup_{\rho \in \mathbb{H}_d(\boldsymbol{\beta}, \mathcal{L})} (\mathbf{E}_b^{(T)}[\varphi_T^{-q}(\boldsymbol{\beta} + \mathbf{1}) \|\widehat{\rho}_{\widehat{\mathbf{h}}} - \rho\|_{\infty}^q])^{1/q} < \infty.$$

REMARK 3.5. The function γ_p° appearing in the definition of α_\circ and Λ is a constant multiple of the function $\gamma_p: \mathbb{R} \times \mathbb{R}_+ \rightarrow \mathbb{R}_+$ introduced in Section 2.3 of [15]. All of our auxiliary results are explicit such that all ingredients for the calculation of γ_p° are available (see in particular Lemma B.1 in Appendix B). Specification of the constants in the definition of γ_p° thus is only a matter of cumbersome, but straightforward computation.

We proceed by establishing an asymptotic lower bound on the sup-norm loss for invariant density estimation under anisotropic Hölder smoothness constraints. In particular, this leads us to considering probability measures of the form $\mu_0 = \exp(-2V_0)/Z_0$, Z_0 denoting the normalising constant. To construct suitable families of hypotheses in the proof of the lower bound, we use a classical transformation (adding/removing a gradient field). The classical Holley–Stroock perturbation lemma states that, if μ_0 satisfies a logarithmic Sobolev inequality, v is a bounded function and $\mu_1 := e^{-v}\mu_0/Z_1$ is another probability measure obtained from μ_0 by multiplication by e^{-v} , then also μ_1 satisfies a logarithmic Sobolev inequality. Even more, it is now classical that all Poincaré–Sobolev-type inequalities are stable under bounded perturbations. This convenient stability property is utilised in the proof of the next result which confirms optimality of the adaptive results stated in Theorem 3.4.

THEOREM 3.6. For any $\beta, \mathcal{L} \in (0, \infty)^d$, $d \geq 3$ and $q \geq 1$,

$$\liminf_{T \rightarrow \infty} \inf_{\hat{\rho}_T} \sup_{\rho \in \mathbb{H}_d(\beta, \mathcal{L})} (\mathbf{E}_b^{(T)} [\varphi_T^{-q}(\beta + \mathbf{1}) \|\hat{\rho}_T - \rho\|_\infty^q])^{1/q} > 0,$$

where the infimum is taken over all possible estimators $\hat{\rho}_T$ of ρ .

Estimation beyond the class of gradient drift diffusions. Let us finally briefly motivate our study of the special case of diffusion processes with unit diffusion part whose drift function is expressed via some potential function. There is good reason to start by investigating estimators based on continuous observations since powerful theoretical results can be proven in this regime. On the one hand, these findings provide benchmarks, showing what ideally could be achieved by discrete-data based methods (see Section 4.1 below). On the other hand, as will be shown in Section 4.2, the proposed procedure can be extended in a straightforward manner to discrete observation schemes.

Having chosen the framework of continuous observations where the diffusion matrix is identifiable, it means only little loss of generality to restrict to the *unit diffusion case*. Furthermore, it is generally very convenient to restrict to *reversible diffusion processes* first since this natural assumption enlightens calculations. It also allows to use a number of nice analytical results for the associated Markov semigroup [e.g., the equivalence between the Poincaré inequality (PI) and the spectral gap inequality (2.1)]. However, according to Kolmogorov’s characterisation of

reversible diffusions, any diffusion process solving (2.2) is reversible with stationary distribution μ on \mathbb{R}^d if and only if there exists a scalar potential $V : \mathbb{R}^d \rightarrow \mathbb{R}$ such that $b = -\nabla V$.

While all of the previous results are stated in this specific framework, inspection of the proof of Theorems 3.3 and 3.4 shows that it is possible to formulate extensions which hold for a substantially larger class of processes. Assuming that \mathbf{X} is reversible and ergodic with invariant density ρ , the bandwidth selection procedure in Section 3.1 allows to define kernel density estimators $\widehat{\rho}_{\mathbf{h}}$ as in (3.8), attaining the rate of convergence $\varphi_T(\boldsymbol{\beta} + \mathbf{1})$, whenever:

- (i) the spectral gap inequality (2.1) is satisfied and
- (ii) a Gaussian bound on the transition densities $p_t(\cdot, \cdot)$ holds *locally*.

Weighted Nash inequalities can be used to prove such nonuniform bounds and may provide an interesting possibility to broaden the scope of a statistical framework incorporating functional inequalities.

4. Estimation from discrete observations. Up to now, the analysis has been conducted for estimators based on continuous observation of the diffusion process during some time period $[0, T]$ under the asymptotics $T \rightarrow \infty$. While this assumption on the observation scheme is highly idealised, the previous results and techniques also prove very useful in the statistical analysis of estimators based on *discrete* observations

$$(4.1) \quad \begin{aligned} X_0 &= X_{T_0}, & X_{T_1}, \dots, X_{T_n}, \\ \Delta_n &:= \max_{i=1, \dots, n} (T_i - T_{i-1}) > 0, & n \in \mathbb{N}, \end{aligned}$$

of a strictly stationary diffusion process $\mathbf{X} = (X_t)_{t \geq 0}$ satisfying the conditions from Definition 1.1.

4.1. *Estimation based on high-frequency observations.* We first consider the context where the data (4.1) are sampled at a high frequency, that is, $\Delta_n \rightarrow 0$ as $n \rightarrow \infty$. In addition, we assume that the total length of the time interval $[0, T_n]$ where observations are made tends to infinity. This is a classical framework for ergodic diffusion models since it allows to estimate the characteristics of the process. Let $K : \mathbb{R} \rightarrow \mathbb{R}$ be as in Theorem 3.4. Inspection of the proof of this theorem shows that, given a continuous record $(X_t)_{0 \leq t \leq T_n}$ of observations and for any $\boldsymbol{\beta} \in (0, \mathfrak{b}]^d$, $\mathcal{L} \in (0, \infty)^d$, the rate-optimal bandwidth choice

$$(4.2) \quad h_j \sim \left(\frac{\log T_n}{T_n} \right)^{\frac{\beta_j + 1 / (\beta_j + 1)}{2\beta_j + 1 + d - 2}}, \quad j = 1, \dots, d,$$

ensures that $\rho_V \in \mathbb{H}_d(\boldsymbol{\beta}, \mathcal{L})$ can be estimated with rate $\varphi_T(\boldsymbol{\beta} + \mathbf{1})$ [cf. (1.12) and (1.13)].

Let us now pass to the case of discrete observations. Define $\mathbb{K}_{\mathbf{h}}$ according to (1.10), and introduce the estimator

$$(4.3) \quad \overset{\Delta}{\rho}_{n,\mathbf{h}}(x) := \frac{1}{T_n} \sum_{i=1}^n \mathbb{K}_{\mathbf{h}}(X_{T_i} - x)(T_i - T_{i-1}), \quad x \in \mathbb{R}^d, \mathbf{h} > 0.$$

Our interest is in specifying a sampling frequency of the form $\Delta_n = O(T_n^{-q})$, for some $q > 0$, such that the convergence rate $\varphi_T = \varphi_{T_n}$ as defined in (1.12) still can be attained.

THEOREM 4.1. *Let $K : \mathbb{R} \rightarrow \mathbb{R}$ be a symmetric, Lipschitz-continuous kernel satisfying (K), and, for some integer $\mathfrak{b} \geq 2$, (3.3). Suppose that \mathbf{X} is strictly stationary, that there exists some constant $M > 0$ such that, for any $x \in \mathbb{R}^d$, $\|b(x)\| \leq M(1 + \|x\|)$, and that $\mathbf{E}_{\mathfrak{b}}[\|X_0\|^2] < \infty$. If the sampling frequency is such that*

$$(4.4) \quad \Delta_n = O\left((\log T_n / T_n)^{\frac{2\beta+1 \cdot \frac{\beta_*+2}{\beta_*+1} + 2d}{2\beta+1+d-2}} \right), \quad \beta_* := \min_{j=1,\dots,d} \beta_j,$$

and the bandwidth $\mathbf{h} = (h_1, \dots, h_d)^\top$ is chosen as in (4.2), the estimator (4.3) satisfies, for any $d \geq 3$, $\beta \in (0, \mathfrak{b}]^d$, $\mathcal{L} \in (0, \infty)^d$,

$$\limsup_{n \rightarrow \infty} \sup_{\rho \in \mathbb{H}_d(\beta, \mathcal{L})} (\mathbf{E}_b^{(T_n)}[\varphi_{T_n}^{-q}(\beta + \mathbf{1}) \|\overset{\Delta}{\rho}_{n,\mathbf{h}} - \rho\|_\infty^q])^{1/q} < \infty.$$

REMARK 4.2. Let us briefly comment on the above result. Most of the findings on (minimal) conditions on the sampling step size Δ_n in high-frequency diffusion models refer to the one-dimensional setup. For the concrete question of nonparametric drift estimation, it is then standard to assume that the number n of sampling points and the sampling frequency Δ_n are such that $n\Delta_n^2 \rightarrow 0$ as $n \rightarrow \infty$. In the isotropic case, the condition (4.4) on the sampling frequency Δ_n for estimating the invariant density of Hölder smoothness $\beta + 1$ from *equidistant* observations $X_0, X_{\Delta_n}, \dots, X_{n\Delta_n}$ with the “fast” rate is fulfilled for

$$\Delta_n \sim \left(\frac{\log n}{n} \right)^{\frac{2\beta+2d+4}{4\beta+3d+4}}.$$

There is a clear dimensional effect: For fixed smoothness β and very large dimension d , the above condition is satisfied for $\Delta_n \sim (\log n/n)^{2/3}$. By way of contrast, for fixed dimension d and increasing smoothness β , the condition on the frequency is close to requiring $\Delta_n \sim (\log n/n)^{1/2}$. For small β , the approximation error $\|\widehat{\rho}_{T_n,\mathbf{h}} - \overset{\Delta}{\rho}_{n,\mathbf{h}}\|_\infty$ between the estimators based on continuous and discrete observations, respectively, can be kept sufficiently small only for a correspondingly small sampling frequency, resulting in a very restrictive condition on Δ_n .

4.2. *An adaptive estimation scheme for low-frequency observations.* Suppose now that \mathbf{X} is observed at equidistant points

$$T_i = i \Delta, \quad i = 0, 1, \dots, n - 1, \Delta > 0 \text{ fixed.}$$

In contrast to the framework of continuous or high-frequency observations, nice path properties of diffusion processes cannot be used in the low-frequency setup. Instead, one has to exploit the Markovian structure.

Given discrete observations $(X_{i\Delta})_{i=0,1,\dots,n-1}$ and some kernel $K : \mathbb{R} \rightarrow \mathbb{R}$, a natural invariant density estimator in the “fixed Δ ”-regime is given by

$$(4.5) \quad \tilde{\rho}_{\mathbf{h}}(x) = \tilde{\rho}_{n,\mathbf{h}}(x) := \frac{1}{n} \sum_{j=0}^{n-1} \mathbb{K}_{\mathbf{h}}(X_{j\Delta} - x), \quad x \in \mathbb{R}^d, \mathbf{h} > 0.$$

Generally speaking, different frequency assumptions require very different investigational approaches and methods. The estimation procedure described in Section 3.1 however is rather “universal” in the sense that its core is *not* related to a specific observation scheme. In particular, it is straightforward to adapt the arguments from the framework of continuous observations and to propose the following bandwidth selection procedure which allows to define estimators of the form (4.5) in a purely data-driven way.

input data:

- * observations $X_0, X_{\Delta}, \dots, X_{(n-1)\Delta}$ of a diffusion process from Definition 1.1, $\Delta > 0$ some fixed observation distance
- * symmetric Lipschitz-continuous kernel $K : \mathbb{R} \rightarrow \mathbb{R}$ with Lipschitz constant $L > 0$ satisfying (K) and, for some integer $\mathfrak{b} \geq 2$, condition (3.3)
- * for $\gamma_p^\bullet(\cdot, \cdot; \cdot)$ defined in Appendix C (cf. the Supplementary Material [26]), set

$$\mathfrak{a}_\bullet := (2\gamma_{2q}^\bullet(d, k_\infty; C_0))^{-2} \quad \text{and} \quad \lambda^\bullet := (1 \vee k_1^d) \gamma_{2q}^\bullet(d, k_\infty; C_0)$$

procedure:

- * introduce the set \mathcal{H}_n of candidate bandwidths by letting

$$\mathcal{H}_n := \left\{ \mathbf{h} = (h_1, \dots, h_d)^\top \in (0, 1]^d : \mathfrak{a}_\bullet^2 \prod_{j=1}^d h_j \geq \frac{\log n}{n} \right\},$$

and denote by $\overline{\mathcal{H}}_n$ a suitable dyadic grid in \mathcal{H}_n

- * define

$$\bar{\varrho}_n := 2 \left(1 \vee \sup_{\mathbf{h} \in \overline{\mathcal{H}}_n} \left\| n^{-1} \sum_{j=0}^{n-1} |\mathbb{K}_{\mathbf{h}}(X_{j\Delta} - \cdot)| \right\|_\infty \right)$$

* given any $\mathbf{h}, \eta \in \overline{\mathcal{H}}_n$, construct estimators $\tilde{\rho}_{\mathbf{h}}$ [cf. (4.5)] and

$$\tilde{\rho}_{\mathbf{h}, \eta}(\cdot) := \frac{1}{n} \sum_{j=0}^{n-1} (\mathbb{K}_{\mathbf{h}} \star \mathbb{K}_{\eta})(X_{j\Delta} - \cdot)$$

* define

$$\begin{aligned} \tilde{A}_n(\mathbf{h}) &:= \sqrt{\bar{q}_n \log n / \left(n \prod_{j=1}^d h_j \right)}, \\ \tilde{\Delta}_n(\mathbf{h}) &:= \sup_{\eta \in \overline{\mathcal{H}}_n} \{ [\|\tilde{\rho}_{\mathbf{h}, \eta} - \tilde{\rho}_{\eta}\|_{\infty} - \lambda \bullet \tilde{A}_n(\eta)]_+ \} \end{aligned}$$

* denote by $\tilde{\rho}_{\tilde{\mathbf{h}}}$ the estimator associated with the bandwidth $\tilde{\mathbf{h}}$ fulfilling

$$(4.6) \quad \tilde{\Delta}_n(\tilde{\mathbf{h}}) + \lambda \bullet \tilde{A}_n(\tilde{\mathbf{h}}) = \min_{\mathbf{h} \in \overline{\mathcal{H}}_n} \{ \tilde{\Delta}_n(\mathbf{h}) + \lambda \bullet \tilde{A}_n(\mathbf{h}) \}$$

The proof of the following result is very close to the proofs of Theorems 3.3 and 3.4. The central modification consists in the treatment of the kernel estimation process which is now based on discrete diffusion observations. For analysing this process, we rely in particular on a Bernstein-type inequality for reversible Markov chains due to [24].

THEOREM 4.3. *Let $K : \mathbb{R} \rightarrow \mathbb{R}$ be a symmetric, Lipschitz-continuous kernel satisfying (K), and, for some integer $\mathfrak{b} \geq 2$, (3.3). If \mathbf{X} is strictly stationary, the adaptive estimator*

$$\tilde{\rho}_{\tilde{\mathbf{h}}}(x) = \frac{1}{n} \sum_{j=0}^{n-1} \mathbb{K}_{\tilde{\mathbf{h}}}(X_{j\Delta} - x), \quad x \in \mathbb{R}^d,$$

with bandwidth $\tilde{\mathbf{h}}$ defined according to (4.6) satisfies, for any $d \geq 1$, $\beta \in (0, \mathfrak{b}]^d$ and any $\mathcal{L} \in (0, \infty)^d$,

$$(4.7) \quad \limsup_{n \rightarrow \infty} \sup_{\rho \in \mathbb{H}_d(\beta, \mathcal{L})} \left(\mathbf{E}_b^{(n)} \left[\left(\frac{\log n}{n} \right)^{-\frac{q\beta+1}{2\beta+1+d}} \|\tilde{\rho}_{\tilde{\mathbf{h}}} - \rho\|_{\infty}^q \right] \right)^{1/q} < \infty.$$

APPENDIX A: COMPLEMENTS ON FUNCTIONAL INEQUALITIES

In recent years, a lot of interest was devoted to the study of (the validity of) functional inequalities and their implications. Examples where Poincaré—and also logarithmic Sobolev—inequalities hold can be obtained under the *Bakry–Émery curvature condition* which involves log-concavity assumptions on the underlying measure. Alternatively, one may use Lyapunov-type criteria which are known to

entail certain results on the long-time behaviour of the laws of Markov processes. The connection between Lyapunov conditions and Poincaré inequalities was studied in detail in [2]. We cite the following concrete conditions for the verification of (PI).

LEMMA A.1 (Corollary 4.8.2 in [3]). *Let $d\mu = e^{-V} d\lambda$ be a probability measure on the Borel sets of \mathbb{R}^d , where $V : \mathbb{R}^d \rightarrow \mathbb{R}$ is a smooth potential such that,*

$$\forall x, \quad \nabla \nabla V(x) \geq \tau \mathbf{Id},$$

for some $\tau > 0$. Then μ satisfies (PI) with $C_0 = \tau^{-1}$.

LEMMA A.2 (Section 4.3 in [2]). *Consider a d -dimensional diffusion process \mathbf{X} solving the SDE $dX_t = -\nabla V(X_t) dt + dW_t$. If there exist M and $r > 0$ such that,*

$$\forall x \text{ with } \|x\| \geq M, \quad \langle \nabla V(x), x \rangle \geq r \|x\|,$$

then \mathbf{X} has a unique invariant probability measure $d\mu = e^{-2V} d\lambda$, and (PI) holds.

The following two auxiliary results play an important role in the proof of the upper and the lower bound results for invariant density estimation. They actually motivate the functional inequalities approach advocated in this paper.

For proving the upper bound, we exploit the well-known relation between the Nash inequality and ultracontractivity (see Section 6.3 in [3] and references therein). Given some diffusion \mathbf{X} with associated Markov semigroup $(P_t)_{t \geq 0}$ and $(p, q) \in [1, \infty)^2$, denote

$$\|P_t\|_{p,q} := \sup \left\{ \|P_t f\|_{L^q(\mu)} : f \in C(\mathbb{R}^d), \|f\|_{L^p(\mu)} = 1, \int f d\mu = 0 \right\}$$

(standard extension to the case $q = \infty$). Classical results are due to [8], showing the equivalence of certain Sobolev-type inequalities and ultracontractivity. Another classical reference is [4]: In the framework of symmetric Markov processes, their Theorem 2.1 asserts that (NI) implies that there exists a constant $B \in (0, \infty)$, depending only on d and C_2 , such that

$$\|P_t\|_{1,\infty} \leq B e^{C_1 t / C_2} t^{-d/2}, \quad t > 0.$$

We shall use the following (related) result.

LEMMA A.3. *Let \mathbf{X} be a diffusion with associated infinitesimal generator L and Markov semigroup $(P_t)_{t \geq 0}$. If L satisfies (NI) with respect to the measure μ with constants $C_1, C_2 > 0$, then*

$$\|P_t\|_{1,\infty} \leq (2C_1 \vee 4dC_2 t^{-1})^{d/2}, \quad t > 0.$$

PROOF. Under the given assumptions, it follows from Theorem 2.1 in [1] that

$$\|P_t f\|_{L^2(\mu)} \leq C(t) \|f\|_{L^1(\mu)} \quad \text{for } C(t) := (2C_1 \vee 2dC_2 t^{-1})^{d/4}.$$

Taking adjoints, we obtain $\|P_t\|_{2,\infty} = \|P_t\|_{1,2}$. The semigroup property implies that $P_t = P_{t/2} \circ P_{t/2}$ such that

$$\|P_t\|_{1,\infty} \leq \|P_{t/2}\|_{1,2} \|P_{t/2}\|_{2,\infty} = \|P_{t/2}\|_{1,2}^2 \leq C^2(t/2). \quad \square$$

LEMMA A.4. *Let \mathbf{X} be a diffusion with associated infinitesimal generator L , Markov semigroup $(P_t)_{t \geq 0}$ and invariant measure μ defined by*

$$d\mu = \left(\int e^{-V} d\lambda \right)^{-1} e^{-V} d\lambda,$$

$V \in C^2(\mathbb{R}^d)$ some smooth potential function with $\int e^{-V} d\lambda < \infty$. Assume that, for $V' \in C^2(\mathbb{R}^d)$, $\|V - V'\|_\infty \leq C$, and consider the probability measure μ' defined by

$$d\mu' = \left(\int e^{-V'} d\lambda \right)^{-1} e^{-V'} d\lambda.$$

(i) *If L satisfies (PI) w.r.t. μ with $C_0 > 0$, then the perturbed process satisfies (PI) w.r.t. μ' with constant $C'_0 \leq e^{-4C} C_0$.*

(ii) *If L satisfies (NI) w.r.t. μ with $C_1, C_2 > 0$, then the perturbed process satisfies (NI) w.r.t. μ' with constants $C'_1 \leq C_1, C'_2 \leq C_2$.*

PROOF. The first assertion is Lemma 1.2 in [14]. For the proof of the second statement, consider the positive and self-adjoint operators

$$A := -L = -\frac{1}{2} \Delta + \nabla V \cdot \nabla \quad \text{and} \quad A' := -\frac{1}{2} \Delta + \nabla V' \cdot \nabla.$$

Trotter's product formula implies that, for $t > 0$,

$$P'_t f := e^{-A't} f = \lim_{n \rightarrow \infty} (e^{-At/n} e^{-\nabla(V'-V)t/n})^n f$$

such that $\|P'_t f\|_\infty \leq \|P_t f\|_\infty$. The assertion follows from the equivalence of ultracontractive estimates and Nash inequalities as in Corollary 2.4.7 in [8]. \square

APPENDIX B: UPPER FUNCTIONS FOR POSITIVE RANDOM FUNCTIONALS IN THE DIFFUSION FRAMEWORK

B.1. Preliminaries. The proposed selection rule for adaptive estimation of the invariant density is based on the computation of *upper functions*. The proof of

the central auxiliary result which we require substantially consists of suitable modifications of the arguments of the proof of Proposition 1 in [15] which itself relies on a rather involved treatment of upper functions. Indeed, the precise implications are as follows:

- Proposition 2 in [16] \implies Theorem 1 in [17]
- \implies Theorem 1 in [18]
- \implies Corollary 1 in [18]
- \implies Theorem 5 in [18]
- \implies Proposition 1 in [15].

The derivation of exact constants is beyond the scope of this paper. We thus content ourselves with sketching the crucial modification in the application of Proposition 2 in [16] (whose conditions can also be verified in the diffusion case) in order to derive a “diffusion specific” version of Theorem 1 in [17].

Let $\mathbf{X} = (X_t)_{t \geq 0}$ be a diffusion process as introduced in Definition 1.1, and assume that \mathbf{X} is strictly stationary. Given some set \mathfrak{H} and a mapping $G : \mathfrak{H} \times \mathbb{R}^d \rightarrow \mathbb{R}$, put

$$\xi_{\mathfrak{h}}(t) := \frac{1}{t} \int_0^t [G(\mathfrak{h}, X_u) - \mathbf{E}_b G(\mathfrak{h}, X_0)] du, \quad \mathfrak{h} \in \mathfrak{H}, t > 0.$$

Throughout the sequel, we will assume that

$$(B.1) \quad \overline{G}_\infty(\mathfrak{h}) := \sup_{x \in \mathbb{R}^d} |G(\mathfrak{h}, x)| < \infty \quad \forall \mathfrak{h} \in \mathfrak{H}.$$

This implies in particular that the random variables $G(\mathfrak{h}, X_0)$, $\mathfrak{h} \in \mathfrak{H}$, and $G(\mathfrak{h}_1, X_u) - G(\mathfrak{h}_2, X_v)$, $\mathfrak{h}_1, \mathfrak{h}_2 \in \mathfrak{H}$, $u, v \geq 0$, are bounded, and we obtain in view of the Bernstein-type inequality (BI) for any $z > 0$ the exponential bound

$$\mathbf{P}_b^{(t)}(|\xi_{\mathfrak{h}}(t)| > z) \leq 2 \exp\left(-\frac{z^2}{A_b^2(\mathfrak{h}) + z B_\infty(\mathfrak{h})}\right),$$

where

$$A_b^2(\mathfrak{h}) := 2t^{-1} \lim_{T \rightarrow \infty} T^{-1} \text{Var}_b\left(\int_0^T G(\mathfrak{h}, X_u) du\right),$$

$$B_\infty(\mathfrak{h}) := 2t^{-1} C_0 \sup_{x \in \mathbb{R}^d} |G(\mathfrak{h}, x)|.$$

The crucial point for finding a “diffusion-tailored” result is to derive an upper bound on $A_b^2(\cdot)$ by combining the variance decay estimate (implied by the spectral gap inequality) with a uniform decay estimate (which follows from Nash inequalities). This is done in the proof of the next auxiliary result.

LEMMA B.1. Assume that the function $G : \mathfrak{H} \times \mathbb{R}^d \rightarrow \mathbb{R}$ satisfies (B.1) and, for any $\mathfrak{h} \in \mathfrak{H}$, $\lambda(\mathfrak{G}_{\mathfrak{h}}) < 1$, where $\mathfrak{G}_{\mathfrak{h}} := \text{supp}(G(\mathfrak{h}, \cdot))$. Then, for any drift function $b = -\nabla V \in \Sigma(C, C_0, C_1, C_2)$ with associated invariant measure $\mu_V = \mu$ and any $\mathfrak{h} \in \mathfrak{H}$, it holds

$$(B.2) \quad A_b^2(\mathfrak{h}) \leq t^{-1} \overline{\mathbf{G}}_{\infty}(\mathfrak{h}) \|G(\mathfrak{h}, \cdot)\|_{L^1(\mu)} \lambda(\mathfrak{G}_{\mathfrak{h}}) \psi_d^2(\lambda(\mathfrak{G}_{\mathfrak{h}}), C_0, C_1, C_2),$$

where, for $x \in (0, 1)$,

$$\begin{aligned} \psi_2^2(x, C_0, C_1, C_2) &:= -4 \log(|x|)(1 - C_0 C_1) + 2C_0 \\ &\quad + 32C_2(\log(-\log(|x|)C_0/2) - \log(-|x| \log(|x|))) \end{aligned}$$

and, for $d \geq 3$,

$$\begin{aligned} \psi_d^2(x, C_0, C_1, C_2) &:= |x|^{(2/d)-1} \left\{ 32dC_2 \left(\frac{2}{d-2} \right)^{2/d} \right. \\ &\quad \left. + 2C_0|x|^{1-(2/d)}(1 - (2C_1)^{d/2} \log(|x|)) \right\}. \end{aligned}$$

PROOF. Let \mathbf{X} be as in Definition 1.1 with invariant measure $\mu_V = \mu$. Fix $\mathfrak{h} \in \mathfrak{H}$, denote $g(x) := G(\mathfrak{h}, x)$, $x \in \mathbb{R}^d$, $g_{\circ} := g - \int g \, d\mu$, and recall the definition of $\sigma^2(\cdot)$ according to (3.11).

Symmetry and invariance of μ imply that

$$(B.3) \quad \sigma^2(G(\mathfrak{h}, \cdot)) = \sigma^2(g) \leq 2 \int_0^{\infty} \langle g_{\circ}, P_s g_{\circ} \rangle_{\mu} \, ds.$$

Analogously to the proof of Proposition 1 in [7], the proof of the upper bound in (B.2) relies on a decomposition of the integral on the right hand side of (B.3). Let $0 < \theta < \Theta < \infty$ to be specified later. For $1 \leq p, q \leq \infty$ with $1/p + 1/q = 1$ and $s \geq 0$, Hölder’s inequality implies that

$$(B.4) \quad \langle g_{\circ}, P_s g_{\circ} \rangle_{\mu} \leq \|g\|_{L^p(\mu)} \|P_s g\|_{L^q(\mu)}.$$

First, let $\mathcal{T}_1 := [0, \theta] \cup [\Theta, \infty)$ and $p = q = 2$, and note that (PI) (precisely, the equivalent spectral gap inequality) entails that

$$\begin{aligned} \int_{\mathcal{T}_1} \langle g_{\circ}, P_s g_{\circ} \rangle_{\mu} \, ds &\leq \int_{\mathcal{T}_1} \|g\|_{L^2(\mu)} \|P_s g\|_{L^2(\mu)} \, ds \\ &\leq \int_{\mathcal{T}_1} e^{-2s/C_0} \|g\|_{L^2(\mu)}^2 \, ds \\ (B.5) \quad &\leq \left(\theta + \frac{1}{2} C_0 e^{-2\Theta/C_0} \right) \overline{\mathbf{G}}_{\infty}(\mathfrak{h}) \|g\|_{L^1(\mu)}. \end{aligned}$$

For $s \in \mathcal{T}_2 := [\theta, \Theta]$, we invoke (NI) and Lemma A.3. Applying (B.4) with $p = 1$, $q = \infty$ first, we obtain

$$\begin{aligned}
 & \int_{\mathcal{T}_2} \langle g_\circ, P_s g_\circ \rangle_\mu ds \\
 \text{(B.6)} \quad & \leq \|g\|_{L^1(\lambda)} \|g\|_{L^1(\mu)} \int_\theta^\Theta \{(2\mathcal{C}_1)^{d/2} + (4d\mathcal{C}_2 s^{-1})^{d/2}\} ds \\
 & \leq \lambda(\mathfrak{G}_h) \bar{\mathbf{G}}_\infty(h) \|g\|_{L^1(\mu)} \left\{ (2\mathcal{C}_1)^{d/2} \Theta + (4d\mathcal{C}_2)^{d/2} \int_\theta^\Theta s^{-d/2} ds \right\}.
 \end{aligned}$$

For $d = 2$, (B.6) implies for any $0 < \theta \leq \Theta$ that

$$\int_{\mathcal{T}_2} \langle g_\circ, P_s g_\circ \rangle_\mu ds \leq \lambda(\mathfrak{G}_h) \bar{\mathbf{G}}_\infty(h) \|g\|_{L^1(\mu)} \{2\mathcal{C}_1 \Theta + 8\mathcal{C}_2 (\log \Theta - \log \theta)\}.$$

Letting $\theta := -\log(\lambda(\mathfrak{G}_h))\lambda(\mathfrak{G}_h)$, $\Theta := -\mathcal{C}_0 \log(\lambda(\mathfrak{G}_h))/2$, we obtain

$$\begin{aligned}
 & \int_0^\infty \langle g_\circ, P_s g_\circ \rangle_\mu ds \\
 & \leq \bar{\mathbf{G}}_\infty(h) \|g\|_{L^1(\mu)} \lambda(\mathfrak{G}_h) \left\{ -\log(\lambda(\mathfrak{G}_h))(1 - \mathcal{C}_0 \mathcal{C}_1) + \frac{\mathcal{C}_0}{2} \right. \\
 & \quad \left. + 8\mathcal{C}_2 \log\left(-\log(\lambda(\mathfrak{G}_h)) \frac{\mathcal{C}_0}{2}\right) - 8\mathcal{C}_2 \log(-\log(\lambda(\mathfrak{G}_h))\lambda(\mathfrak{G}_h)) \right\}.
 \end{aligned}$$

The assertion for $d = 2$ follows by substituting this upper bound into

$$A_b^2(h) \leq 4t^{-1} \int_0^\infty \langle g_\circ, P_s g_\circ \rangle_\mu ds.$$

For $d \geq 3$, it holds in view of (B.5) and (B.6),

$$\begin{aligned}
 \int_0^\infty \langle g_\circ, P_s g_\circ \rangle_\mu ds & \leq \left\{ \theta + \frac{1}{2} \mathcal{C}_0 e^{-2\Theta/\mathcal{C}_0} + \lambda(\mathfrak{G}_h) (2\mathcal{C}_1)^{d/2} \Theta \right. \\
 & \quad \left. + \frac{2(4d\mathcal{C}_2)^{d/2}}{d-2} \theta^{1-(d/2)} \lambda(\mathfrak{G}_h) \right\} \bar{\mathbf{G}}_\infty(h) \|g\|_{L^1(\mu)}.
 \end{aligned}$$

Set $\theta := 4d\mathcal{C}_2(2/(d-2))^{2/d} \lambda(\mathfrak{G}_h)^{2/d}$, $\Theta := -\mathcal{C}_0 \log(\lambda(\mathfrak{G}_h))/2$. Then

$$\begin{aligned}
 \int_0^\infty \langle g_\circ, P_s g_\circ \rangle_\mu ds & \leq \lambda(\mathfrak{G}_h) \bar{\mathbf{G}}_\infty(h) \|g\|_{L^1(\mu)} \left\{ 8d\mathcal{C}_2 \left(\frac{2}{d-2}\right)^{2/d} \lambda(\mathfrak{G}_h)^{(2/d)-1} \right. \\
 & \quad \left. + \frac{\mathcal{C}_0}{2} (1 - (2\mathcal{C}_1)^{d/2} \log(\lambda(\mathfrak{G}_h))) \right\},
 \end{aligned}$$

and the claim follows as above. \square

B.2. Upper functions on the kernel estimation process. Consider some function $M : \mathbb{R} \rightarrow \mathbb{R}$ satisfying the following assumptions:

$$\int M \, d\lambda = 1, \quad m_\infty := \|M\|_\infty < \infty, \quad \text{supp}(M) \subset [-1/2, 1/2],$$

M is symmetric and $\exists L > 0$ such that $|M(t) - M(s)| \leq L|t - s| \, \forall t, s \in \mathbb{R}$. For any $\mathbf{r} \in (0, 1]^d, x = (x_1, \dots, x_d)^\top \in \mathbb{R}^d$ and $V_{\mathbf{r}}$ defined according to (1.10), set

$$\mathbb{M}_{\mathbf{r}}(x) := V_{\mathbf{r}}^{-1} \prod_{l=1}^d M(x_l/r_l) = \prod_{l=1}^d r_l^{-1} M(x_l/r_l).$$

Introduce families of random fields by setting

$$\chi_{\mathbf{r}}(x) := \frac{1}{T} \int_0^T [\mathbb{M}_{\mathbf{r}}(X_u - x) - \mathbf{E}_b^{(T)}[\mathbb{M}_{\mathbf{r}}(X_0 - x)]] \, du, \quad x \in \mathbb{R}^d,$$

with $\mathbf{r} \in \tilde{\mathcal{R}}_T(d) := \{\mathbf{r} \in (0, 1]^d : T \geq \psi_d^2(V_{\mathbf{r}}) \log T\}$,

and

$$Z_{\mathbf{r}}(x) := \frac{1}{T} \int_0^T |\mathbb{M}_{\mathbf{r}}(X_u - x)| \, du, \quad x \in \mathbb{R}^d,$$

with $\mathbf{r} \in \tilde{\mathcal{R}}_T^{(a)}(d) := \{\mathbf{r} \in (0, 1]^d : T a^2 \geq \psi_d^2(V_{\mathbf{r}}) \log T\}$,

$a := (2\gamma_p^\circ(d, m_\infty))^{-2}$ for the constant $\gamma_p^\circ(d, m_\infty) \equiv \gamma_p^\circ(d, m_\infty; C_0, C_1, C_2)$ described in Remark 3.5. For any $\mathbf{r} \in (0, 1]^d, y \in \mathbb{R}^d$, define

$$G(\mathbf{r}, y) := \int_{\mathbb{R}^d} |\mathbb{M}_{\mathbf{r}}(x - y)| \rho_V(x) \, dx = \|\mathbb{M}_{\mathbf{r}}(\cdot - y)\|_{L^1(\mu)},$$

$G(\mathbf{r}) := \sup_{y \in \mathbb{R}^d} G(\mathbf{r}, y)$ and $\overline{G}(\mathbf{r}) := 1 \vee G(\mathbf{r})$.

PROPOSITION B.2 (cf. Propositions 1 and 2 in [15]). *Let \mathbf{X} be as in Definition 1.1. Then, for any $T > 0$ and any $p \geq 1$, there exist constants c_1, c_2, c'_1, c'_2 , depending only on $p, d, \|M\|_{L^1(\lambda)}, m_\infty$, such that*

$$\begin{aligned} \text{(B.7)} \quad & \mathbf{E}_b^{(T)} \left[\sup_{\mathbf{r} \in \tilde{\mathcal{R}}_T(d)} \left\{ \|\chi_{\mathbf{r}}\|_\infty - \gamma_p^\circ(d, m_\infty) \psi_d(V_{\mathbf{r}}) \sqrt{\frac{\overline{G}(\mathbf{r}) \log T}{T}} \right\} \right]_+^p \\ & \leq c_1 T^{-p/2} + c_2 T^{-p}, \end{aligned}$$

$$\text{(B.8)} \quad \mathbf{E}_b^{(T)} \left[\sup_{\mathbf{r} \in \tilde{\mathcal{R}}_T^{(a)}(d)} \left\{ \overline{G}(\mathbf{r}) - 2(1 \vee \|Z_{\mathbf{r}}\|_\infty) \right\} \right]_+^p \leq c'_1 T^{-p/2} + c'_2 T^{-p},$$

$$\text{(B.9)} \quad \mathbf{E}_b^{(T)} \left[\sup_{\mathbf{r} \in \tilde{\mathcal{R}}_T^{(a)}(d)} \left\{ 1 \vee \|Z_{\mathbf{r}}\|_\infty - \frac{3}{2} \overline{G}(\mathbf{r}) \right\} \right]_+^p \leq c_1 T^{-p/2} + c_2 T^{-p}.$$

PROOF. Under the above conditions, Assumptions 3, 5, 8 in Theorem 5 in [18] continue to hold in the diffusion framework. Obviously, $\lambda(\text{supp}(\mathbb{M}_{\mathbf{r}})) \sim V_{\mathbf{r}}$. In view of

$$\sup_{x \in \mathbb{R}^d} |\mathbb{M}_{\mathbf{r}}(x)| \leq V_{\mathbf{r}}^{-1} m_{\infty}^d,$$

Lemma B.1 thus implies that

$$A_b^2(\mathbf{r}) := 2T^{-1} \sigma^2(\mathbb{M}_{\mathbf{r}}) \leq C_d(C_0, C_1, C_2) m_{\infty}^d T^{-1} \psi_d^2(V_{\mathbf{r}}) \overline{G}(\mathbf{r}).$$

This observation allows to verify that a correspondent of the upper function $\widehat{U}^{(v,z,p)}$ in Theorem 5 in [18] (with matched notation) is given by the function $\overline{U}^{(v,z,p)}(T, \mathbf{r}, \overline{x}^{(d)})$, specified as

$$\begin{aligned} \lambda_1^{\circ} \sqrt{\frac{\psi_d^2(V_{\mathbf{r}}) \overline{G}(\mathbf{r})}{T} [\widehat{M}_{q,v}(\mathbf{r}) + 2(v+1) |\log(\max\{G(\mathbf{r}, \overline{x}^{(d)}), T^{-1}\})| + z]} \\ + \frac{\lambda_2^{\circ}}{TV_{\mathbf{r}}} [\widehat{M}_{q,v}(\mathbf{r}) + 2(v+1) |\log(\max\{G(\mathbf{r}, \overline{x}^{(d)}), T^{-1}\})| + z], \end{aligned}$$

where $\lambda_1^{\circ} = \lambda_1^{\circ}(C_0, C_1, C_2)$, $\lambda_2^{\circ} = \lambda_2^{\circ}(C_0)$ are constant multiples of the quantities λ_1, λ_2 defined in Section 3.1 of [17] (see page 200), and $\widehat{M}_{q,v}(\cdot)$ is defined as in Section 4.4.2 in [15].

In particular, letting $q = p$, $v = 2p + 2$, $z = 1$ and $\overline{x}^{(d)} = y$, we obtain for any $\mathbf{r} \in \widetilde{\mathcal{R}}_T(d) \subset \mathcal{R}_T(d)$ that

$$\overline{U}^{(v,z,p)}(T, \mathbf{r}, \overline{x}^{(d)}) \leq \gamma_p^{\circ}(d, m_{\infty}) \psi_d(V_{\mathbf{r}}) \sqrt{\frac{\overline{G}(\mathbf{r}) \log T}{T}}.$$

A version of Theorem 5 in [18] then immediately gives (B.7). The remaining assertions are deduced from (B.7) precisely as in the proof of Proposition 2 in [15]. □

Acknowledgements. It is a pleasure to thank Patrick Cattiaux and Alexei Kulik for very helpful advice. Further thanks goes to the referees for a very careful reading of the manuscript. Their comments and suggestions led to a number of improvements and stimulated further research.

SUPPLEMENTARY MATERIAL

Supplement to “Adaptive invariant density estimation for ergodic diffusions over anisotropic classes” (DOI: [10.1214/17-AOS1664SUPP](https://doi.org/10.1214/17-AOS1664SUPP); .pdf). The supplementary file contains the proofs of the sup-norm oracle-type inequality stated in Theorem 3.3, the upper bound in Theorem 3.4, the lower bound from Theorem 3.6 and of the results for estimators based on discrete observations stated in Theorem 4.1 and Theorem 4.3, respectively.

REFERENCES

- [1] BAKRY, D., BOLLEY, F. and GENTIL, I. (2011). Around Nash inequalities. In *Journées Équations aux Dérivées Partielles* (2010), Biarritz, France.
- [2] BAKRY, D., CATTIAUX, P. and GUILLIN, A. (2008). Rate of convergence for ergodic continuous Markov processes: Lyapunov versus Poincaré. *J. Funct. Anal.* **254** 727–759. [MR2381160](#)
- [3] BAKRY, D., GENTIL, I. and LEDOUX, M. (2014). *Analysis and Geometry of Markov Diffusion Operators. Grundlehren der Mathematischen Wissenschaften* **348**. Springer, Cham.
- [4] CARLEN, E. A., KUSUOKA, S. and STROOCK, D. W. (1987). Upper bounds for symmetric Markov transition functions. *Ann. Inst. Henri Poincaré Probab. Stat.* **23** 245–287.
- [5] CATTIAUX, P. and GUILLIN, A. (2008). Deviation bounds for additive functionals of Markov processes. *ESAIM Probab. Stat.* **12** 12–29.
- [6] CHOROWSKI, J. and TRABS, M. (2016). Spectral estimation for diffusions with random sampling times. *Stochastic Process. Appl.* **126** 2976–3008. [MR3542623](#)
- [7] DALALYAN, A. and REISS, M. (2007). Asymptotic statistical equivalence for ergodic diffusions: The multidimensional case. *Probab. Theory Related Fields* **137** 25–47.
- [8] DAVIES, E. B. (1990). *Heat Kernels and Spectral Theory. Cambridge Tracts in Mathematics* **92**. Cambridge Univ. Press, Cambridge.
- [9] GOBET, E., HOFFMANN, M. and REISS, M. (2004). Nonparametric estimation of scalar diffusions based on low frequency data. *Ann. Statist.* **32** 2223–2253. [MR2102509](#)
- [10] GOLDENSHLUGER, A. and LEPSKI, O. (2008). Universal pointwise selection rule in multivariate function estimation. *Bernoulli* **14** 1150–1190.
- [11] GOLDENSHLUGER, A. and LEPSKI, O. (2009). Structural adaptation via L_p -norm inequalities. *Probab. Theory Related Fields* **143** 41–71.
- [12] GOLDENSHLUGER, A. and LEPSKI, O. (2011). Bandwidth selection in kernel density estimation: Oracle inequalities and adaptive minimax optimality. *Ann. Statist.* **39** 1608–1632. [MR2850214](#)
- [13] KUTOYANTS, Y. A. (2004). *Statistical Inference for Ergodic Diffusion Processes*. Springer, New York.
- [14] LEDOUX, M. (2001). Logarithmic Sobolev inequalities for unbounded spin systems revisited. In *Séminaire de Probabilités XXXV. Lecture Notes in Math.* **1755** 167–194. Springer, Berlin.
- [15] LEPSKI, O. (2013). Multivariate density estimation under sup-norm loss: Oracle approach, adaptation and independence structure. *Ann. Statist.* **41** 1005–1034. [MR3099129](#)
- [16] LEPSKI, O. (2013). Upper functions for positive random functionals. I. General setting and Gaussian random functions. *Math. Methods Statist.* **22** 1–27. [MR3040410](#)
- [17] LEPSKI, O. (2013). Upper functions for positive random functionals. II. Application to the empirical processes theory, Part 1. *Math. Methods Statist.* **22** 83–99. [MR3071956](#)
- [18] LEPSKI, O. (2013). Upper functions for positive random functionals. II. Application to the empirical processes theory, Part 2. *Math. Methods Statist.* **22** 193–212. [MR3107668](#)
- [19] LEPSKI, O. V. and LEVIT, B. Y. (1999). Adaptive nonparametric estimation of smooth multivariate functions. *Math. Methods Statist.* **8** 344–370.
- [20] LEZAUD, P. (2001). Chernoff and Berry–Esséen inequalities for Markov processes. *ESAIM Probab. Stat.* **5** 183–201.
- [21] NASH, J. (1958). Continuity of solutions of parabolic and elliptic equations. *Amer. J. Math.* **80** 931–954.
- [22] NELSON, E. (1958). The adjoint Markoff process. *Duke Math. J.* **25** 671–690.
- [23] NIKOLSKII, S. M. (1975). *Approximation of Functions of Several Variables and Imbedding Theorems. Die Grundlehren der Mathematischen Wissenschaften* **205**. Springer, New York. Translated from the Russian by J. M. Danskin, Jr. [MR0374877](#)

- [24] PAULIN, D. (2015). Concentration inequalities for Markov chains by Marton couplings and spectral methods. *Electron. J. Probab.* **20** 1–32. [MR3383563](#)
- [25] SÖHL, J. and TRABS, M. (2016). Adaptive confidence bands for Markov chains and diffusions: Estimating the invariant measure and the drift. *ESAIM Probab. Stat.* **20** 432–462. [MR3581829](#)
- [26] STRAUCH, C. (2018). Supplement to “Adaptive invariant density estimation for ergodic diffusions over anisotropic classes.” DOI:[10.1214/17-AOS1664SUPP](#).
- [27] VAN DER VAART, A. W. and VAN ZANTEN, H. (2005). Donsker theorems for diffusions: Necessary and sufficient conditions. *Ann. Probab.* **33** 1422–1451. [MR2150194](#)

SCHOOL OF BUSINESS INFORMATICS AND MATHEMATICS
UNIVERSITY OF MANNHEIM
68131 MANNHEIM
GERMANY
E-MAIL: strauch@uni-mannheim.de