Probab. Theory Relat. Fields (2016) 164:361–400
DOI 10.1007/s00440-014-0614-4

Exact adaptive pointwise drift estimation for multidimensional ergodic diffusions

Claudia Strauch

Received: 31 October 2013 / Revised: 21 December 2014 / Published online: 12 February 2015

Abstract The problem of pointwise adaptive estimation of the drift coefficient of a multivariate diffusion process is investigated. We propose an estimator which is sharp adaptive on scales of Sobolev smoothness classes. The analysis of the exact risk asymptotics allows to identify the impact of the dimension and other influencing values—such as the geometry of the diffusion coefficient—of the prototypical drift estimation problem for a large class of multidimensional diffusion processes. We further sketch generalizations of our results to arbitrary diffusions satisfying suitable Bernstein-type inequalities.

Keywords Ergodic diffusion · Minimax drift estimation · Exact constants in nonparametric smoothing · Sharp adaptivity · Pointwise risk

Mathematics Subject Classification 62M05 · 62G07 · 62G20

1 Introduction and motivation

Diffusions present a particularly important class of stochastic processes. The long standing probabilistic interest in this subject is illustrated, for example, by the seminal books of Itō and McKean [9] and Stroock and Varadhan [22]. From the statistical point of view, one classical problem is to estimate the (unknown) characteristics of the diffusion, both from continuous-time and discrete observations. In the last two decades, nonparametric estimation of diffusion processes has been widely developed, mainly due to their applications in mathematical finance where diffusions are commonly used to model the evolution of financial assets or portfolios of assets. While

C. Strauch (✉)
Universität Hamburg, Bundesstr. 55, 20146 Hamburg, Germany
e-mail: strauch@math.uni-hamburg.de

diffusion models have been largely univariate in the past, they now predominantly include multiple state variables; see the introductory remarks of Aït-Sahalia [1] for concrete examples. In some respects, the statistical theory did not keep pace with this evolution: thorough theoretical results on nonparametric estimation of *multidimensional* diffusion processes are few and far between. At least partially, this is due to the fact that the concept of diffusion local time and related tools such as the occupation times formula are not available in dimension $d > 1$ such that the treatment of the multivariate case requires different approaches and new techniques.

The aim of this paper is to close one gap in the literature by analyzing the asymptotically exact behavior of the pointwise risk for adaptively estimating the drift vector $b : \mathbb{R}^d \to \mathbb{R}^d$ of a multivariate diffusion which is given as a solution of the stochastic differential equation

$$\mathrm{d}X_t = b(X_t)\mathrm{d}t + \sigma(X_t)\mathrm{d}W_t, \quad X_0 = \xi,\ t \in [0, T], \tag{1.1}$$

where $\sigma : \mathbb{R}^d \to \mathbb{R}^{d\times d}$ is the dispersion matrix, W is a d-dimensional standard Wiener process and the initial value $\xi \in \mathbb{R}^d$ is independent of W. It will be assumed throughout that a continuous record of observations $X^T := (X_t)_{0\le t\le T}$ is available. Thus, the diffusion coefficient $\sigma\sigma^\top$ is identifiable by means of the semimartingale quadratic variation, and it means only little loss of generality to simplify the analysis by considering merely the case of known, constant diffusion part. We further restrict attention to ergodic diffusions whose invariant measure is absolutely continuous with respect to the Lebesgue measure. Let ρ_b denote the invariant density. The initial value ξ is assumed to follow the invariant law such that the process X is strictly stationary.

It is statistical folklore to consider drift estimation as some analogue of the regression problem. Given some appropriately chosen kernel $K : \mathbb{R}^d \to \mathbb{R}$ and bandwidth $h > 0$, it thus appears natural to investigate the following type of kernel estimators of the drift vector b,

$$\widehat{b}_T(x) := \frac{T^{-1}\int_0^T K_h(X_u - x)\mathrm{d}X_u}{\widehat{\rho}_T(x) \vee \rho_*(x)}, \quad x \in \mathbb{R}^d, \tag{1.2}$$

where $K_h(\cdot) := h^{-d}K(\cdot/h)$, $\widehat{\rho}_T$ is some estimator of the invariant density ρ_b and $\rho_*(x) > 0$ denotes some a priori lower bound on $\rho_b(x)$. In the sequel, the quality of an estimator $\widehat{b}_T$ will be quantified by its pointwise risk

$$\mathcal{R}\big(\widehat{b}_T, b\big) := \mathbf{E}_b\|\widehat{b}_T(x_0) - b(x_0)\|^2, \quad x_0 \in \mathbb{R}^d \text{ fixed},$$

for $\mathbf{E}_b$ denoting expectation with respect to the invariant measure associated with b and $\|\cdot\|$ denoting the Euclidean norm. The goal is to define minimax adaptive estimators b_T^* satisfying

$$\mathcal{R}\big(b_T^*, b\big) = \inf_{\widetilde{b}_T}\sup_{b\in\mathscr{B}} \mathcal{R}\big(\widetilde{b}_T, b\big).$$

The supremum here extends over a given class of functions $\mathscr{B}$, typically, a class of functions satisfying certain smoothness assumptions or structural constraints. For estimating the drift vector, we shall consider scales of Sobolev classes $(\Sigma_T(\beta, L))_{(\beta,L)\in\mathcal{B}_T}$ where, for fixed $\beta_* > d/2$ and $0 < L_* < L^* < \infty$,

$$\mathcal{B}_T := \left\{(\beta, L) : \beta_* \leq \beta < \beta_T,\ L_* \leq L \leq L^*\right\}, \quad \beta_T = (\log\log T)^{\delta},\ \delta \in (0, 1) \text{ fixed.}$$

We propose estimators which attain not only the optimal rate of convergence but the best *exact* asymptotic minimax risk when the actual smoothness of the drift and the associated invariant density ρ_b are unidentified and we only assume membership to $\Sigma_T(\beta, L)$ for *some* $(\beta, L) \in \mathcal{B}_T$.

To the best of our knowledge, sharp asymptotic minimax bounds for nonparametric estimation in the diffusion framework have been established exclusively for one-dimensional processes up to now. One particularly deep result is given in Dalalyan [5] where a fully data-driven procedure for exact global estimation of the drift for a large class of ergodic scalar diffusion processes is proven. In the multidimensional diffusion set-up however, we only know of upper bound results on *rates* of convergence, even for the prototypical problem of estimating the drift vector from continuous-time observations. Let us emphasize that the question of identifying the exact constant in the risk asymptotics is far from being merely of theoretical interest. The subsequent in-depth analysis rather allows to descry the influencing values of the drift estimation problem, and these findings provide answers to practice-oriented issues. For instance, it is to be expected—and has been observed in practice indeed—that the speed of convergence for estimating functionals of a diffusion process solution of the SDE (1.1) depends on the geometry of the diffusion coefficient $\sigma\sigma^\top =: a = (a_{jk})_{1\leq j,k\leq d}$. For the exemplary problem of estimating the j-th component b^j of the drift vector, $j \in \{1, \ldots, d\}$ fixed, the dependence will be proven to be reflected by the appearance of a_{jj}, the j-th diagonal entry of the diffusion matrix, in the exact normalizing factor in the risk asymptotics. Our exact results further give a theoretical justification for the wide-spread use of standard kernel methods for drift estimation which in applications (e.g., in financial econometrics) often occurs on an *ad-hoc* basis. Heuristically, the use of such methods is based on the aforementioned folklore that "drift estimation is just regression," provided that the long observation limit is considered and as long as the diffusion is sufficiently regular.

On a mathematically formal level, abstract decision theory allows to transfer risk bounds from one statistical model to another by referring to the concept of asymptotic equivalence of experiments in the sense of Le Cam. For inference on the drift in multidimensional ergodic diffusion models, asymptotic equivalence is established in Dalalyan and Reiß [7]. Their results concern the special case of Kolmogorov diffusions with unit diffusion part, i.e. $\sigma = \mathbf{Id}$, and hold only for large enough Hölder smoothness of the drift coefficient (which is substantially larger than the lower bound of $d/2$ which would correspond to known results on asymptotic nonequivalence of scalar nonparametric experiments when the smoothness index is $1/2$). We take a direct approach and establish upper and lower asymptotic risk bounds for diffusions with general constant and nondegenerate diffusion part without resorting to arguments based on asymptotic equivalence.

For ease of presentation however, let us merely announce the result for the important special case of diffusions with dispersion matrix of the form $\sigma = \sigma_0 \, \mathbf{Id}$, for some $0 \neq \sigma_0 \in \mathbb{R}$. Define

$$D(\beta, L; \rho_b, \sigma_0) := \frac{2\beta L^{\frac{d}{2\beta}}}{\rho_b(x_0)\sqrt{d}} \left(\frac{d^2 \sigma_0^2 \rho_b(x_0)}{\beta(2\beta - d)} \right)^{\frac{\beta - d/2}{2\beta}} \mathbb{I}_\beta, \quad \beta > \frac{d}{2}, \ L > 0. \tag{1.3}$$

Here, with $B(\cdot,\cdot)$ and $\Gamma(\cdot)$ denoting the Beta and the Gamma function, respectively, and letting $\mathbb{S}_d := 2\pi^{d/2}/\Gamma(d/2)$ denote the surface of the unit sphere in $\mathbb{R}^d$,

$$\begin{aligned} \mathbb{I}_\beta^2 &:= \frac{1}{2\beta} \, B\left(1 + \frac{d}{2\beta}, \ 1 - \frac{d}{2\beta}\right) (2\pi)^{-d} \mathbb{S}_d \\ &= \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \frac{\|\lambda\|^{2\beta}}{\left(1 + \|\lambda\|^{2\beta}\right)^2} \, \mathrm{d}\lambda. \end{aligned} \tag{1.4}$$

On the one hand, we show that

$$\begin{aligned} \liminf_{T\to\infty} \inf_{\widetilde{b}_T} \sup_{(\beta,L)\in\mathcal{B}_T} \sup_{b\in\Pi(c_1,c_2)} \sup_{\rho_b\in\Sigma_T(\beta,L)} \left(\frac{T}{\log T}\right)^{\frac{\beta-d/2}{\beta}} \\ \times D^{-2}(\beta, L; \rho_b, \sigma_0) \, \mathbf{E}_b \left\| \widetilde{b}_T(x_0) - b(x_0) \right\|^2 \geq 1, \end{aligned}$$

for some suitably defined functional sets $\Pi(c_1, c_2)$ and $\Sigma_T(\beta, L) = \Sigma_T(\beta, L; L', \sigma_0)$, depending also on σ_0 and constants c_1, c_2, L' related to the regularity properties of the class of investigated multivariate diffusion processes (for details, see Sects. 2 and 5). Furthermore, we suggest an asymptotically *sharp adaptive* estimator over $\mathcal{B}_T$, i.e. an adaptive estimator which does not only achieve the best possible rate of convergence but the best asymptotic constant associated to it. Our exact asymptotic results on drift estimation hold under mild regularity constraints and indicate that asymptotic equivalence—at least in some reduced sense—also holds under smoothness assumptions less severe than those imposed in Dalalyan and Reiß [7]; cf. the discussion in Sect. 6.

The current investigation is directly related to previous work both from the field of nonparametric statistics and more applied areas such as financial econometrics. A larger quantity of results on nonparametric drift estimation in the scalar diffusion case is already available. Dalalyan and Kutoyants [6] consider the problem of nonparametric estimation of the derivative of the invariant density and of the drift coefficient for scalar ergodic diffusion processes over weighted L^2 Sobolev classes. The construction of the suggested asymptotically efficient estimator requires the knowledge of the smoothness and the radius of these weighted Sobolev balls. On the basis of these results, Dalalyan [5] develops an adaptive procedure which does not depend on the characteristics of the Sobolev ball and which is asymptotically minimax simultaneously over a broad scale of Sobolev classes. In direct relation to the present work, Spokoiny [20] considers the problem of pointwise adaptive drift estimation and develops a locally linear smoother with data-driven bandwidth choice. His method is also derived in a scalar setting but generalizes to the multidimensional framework. The focus of Spokoiny [20] clearly differs from ours: he provides nonasymptotic results (which do not require stationarity, ergodicity or mixing properties of the observed diffusion process) for the suggested kernel type estimators, while our interest is in identifying the asymptotically exact behavior of adaptive drift estimators. The definition of such asymptotically sharp adaptive estimators does not only require a data-dependent choice of the smoothing parameter but also a data-driven selection of the kernel.

In the sequel, we will use rather recent results on functional inequalities (and the interplay of different types thereof) for diffusion processes. To be more precise, inspection of the constructive proof of the asymptotic upper risk bound suggests that the combination of a Bernstein-type deviation inequality and sufficiently tight variance bounds is the key for suggesting *sharp* adaptive drift estimation procedures for diffusion processes. Diverse works on generalizations of the classical Bernstein inequality which are applicable in the diffusion framework exist. In this paper, we will assume that the diffusion satisfies the spectral gap inequality—a condition which, at least in the area of statistics for random processes, is rather unconventional. However, it can be argued that this hypothesis presents some sort of minimal assumption for a Bernstein-type inequality for symmetric diffusion processes to hold and thus provides a natural framework for our investigation. The combination of different types of tail estimates of additive functionals and sharp variance bounds due to Dalalyan and Reiß [7] then allows to prove the required type of exponential inequalities, and classical chaining arguments and conditions on the size of function classes in terms of bracketing numbers provide an extension to *uniform* versions thereof. Our results still are by no means restricted to this specific kind of dependence mechanism as will be sketched later. Currently, (probabilistic) research on diffusion processes is aimed at investigating the interplay between different approaches for the study of quantitative ergodic properties and the relationship between different functional inequalities. It would be interesting to complement these results with findings on the asymptotic statistical behavior of estimators in the respective ergodic diffusion models, and the present analysis provides one first step in this direction.

Outline of the paper One crucial point in our subsequent investigation is the fact that we may restrict attention to analyzing the exact asymptotics of the estimators which appear in the numerator of (1.2). Only mild regularity properties of the diffusion are required for translating results on estimating

$$\begin{aligned}\lim_{h\to 0}\mathbf{E}_b\left[\frac{1}{T}\int_0^T K_h(X_u-x)\mathrm{d}X_u\right]&=\lim_{h\to 0}\int_{\mathbb{R}^d}K_h(y-x)b(y)\rho_b(y)\mathrm{d}y\\&=b(x)\rho_b(x),\quad x\in\mathbb{R}^d,\end{aligned}$$

into upper and lower bounds for drift estimation. We thus start our investigation with considering estimation of $b\rho_b$, assuming that the components $b^j\rho_b$, $j\in\{1,\dots,d\}$, belong to some Sobolev class of regularity $\beta\in\mathcal{I}$, $\mathcal{I}$ some given interval of the form $[\beta_*,\beta_T]$ with $\beta_T\to_{T\to\infty}\infty$ slowly enough. Section 3 contains a lower bound for pointwise estimation of $b\rho_b$, and an adaptive procedure for estimating the components of $b\rho_b$ which asymptotically attains the respective infimum is introduced in Sect. 4. Provided that the drift grows at most linearly and the invariant density decays exponentially, upper and lower bounds for estimating $b\rho_b$ can be translated into corresponding results for drift estimation. In favor of a concise and transparent presentation, the bounds are stated explicitly only for Kolmogorov diffusions. The respective results are given in Sect. 5. Section 6 contains a discussion of our findings and a sketch of possible extensions. Details on the exponential inequality used in the proof of the upper bound part of our exact result are given in Appendix A. The bulk of the proofs of the main results is deferred to Appendix B.

General definitions and notation For $g:\mathbb{R}^d\to\mathbb{R}^d$, denote by g^j its j-th component. For a smooth function $f:\mathbb{R}^d\to\mathbb{R}$, let $\partial_j f := \partial f/\partial x^j$, and denote its gradient by $\nabla f = (\partial_j f)_j$. Rows of an $d\times d$-matrix a are denoted by a_j, and the Frobenius norm of the matrix σ is denoted by $\|\sigma\|_{S_2} := (\sum_{j=1}^d (\sigma\sigma^\top)_{jj})^{1/2}$. Let ϕ_f be the Fourier transform of $f\in L^2(\mathbb{R}^d)$, that is, for any $\lambda\in\mathbb{R}^d$, $\phi_f(\lambda) := \int_{\mathbb{R}^d} f(x)\exp(\mathrm{i}\,\lambda^\top x)\mathrm{d}x$. Let $\beta > d/2$, and define the Sobolev seminorm $\eta_\beta(\cdot)$ by

$$\eta_\beta(f) := \left(\frac{1}{(2\pi)^d}\int_{\mathbb{R}^d}\|\lambda\|^{2\beta}\left|\phi_f(\lambda)\right|^2\mathrm{d}\lambda\right)^{1/2},\quad f\in L^2(\mathbb{R}^d).$$

The isotropic Sobolev class $\mathcal{S}(\beta,L)$ is given as $\mathcal{S}(\beta,L):=\left\{f\in L^2(\mathbb{R}^d):\eta_\beta(f)\le L\right\}$. Throughout, $\lesssim$ means less or equal up to some constant which does not depend on the variable parameters in the expression.

2 Preliminaries

The complexity of the diffusion model requires some care in defining the framework for pointwise estimation of the components of the drift vector, with special consideration of the interplay of regularity properties of the individual components of the model. We start by defining $\Pi_0 = \Pi_0(\sigma)$, σ some constant nondegenerate $\mathbb{R}^{d\times d}$-valued dispersion matrix with associated diffusion coefficient $\sigma\sigma^\top = a$, as the set of all drift coefficients $b\colon\mathbb{R}^d\to\mathbb{R}^d$ such that

(P_0) the SDE

$$\mathrm{d}X_t = b(X_t)\mathrm{d}t + \sigma\,\mathrm{d}W_t \tag{2.1}$$

admits a strong solution which is ergodic with Lebesgue continuous invariant measure $\mathrm{d}\mu_b(x) = \rho_b(x)\mathrm{d}x$, and

(P_0') for $j\in\{1,\ldots,d\}$, the invariant density ρ_b satisfies the relation

$$2b^j\rho_b = a_j\nabla\rho_b = \sum_{k=1}^d a_{jk}\partial_k\rho_b.$$

We further suppose that the initial value X_0 has the density ρ_b such that $(X_t)_{t\geq 0}$ is strictly stationary.

As aforementioned, the drift estimation problem in the sequel will be decomposed into the individual questions of estimating the invariant density ρ_b and the products $b^j \rho_b, j = 1, \ldots, d$. Restricting to diffusion processes satisfying (P_0'), the second question can also be stated as estimating the weighted sums of derivatives $\sum_{k=1}^d a_{jk} \partial_k \rho_b$, $j = 1, \ldots, d$. As has been proved in Dalalyan and Kutoyants [6] and Dalalyan [5] in the scalar set-up, this approach has the potential to derive deep results. We already noted that the non-existence of diffusion local time presents a particular challenge for the statistical analysis of estimators in the multivariate diffusion framework as a set of valuable technical tools falls away. One further difficulty consists in identifying regularity conditions on the diffusion which allow for an as broad as possible extension of the investigation to a multivariate framework. It is convenient to include the condition (P_0'), but our results can also be generalized to more general classes of diffusion processes.

In the sequel, we consider estimation of the drift function at a point $x_0 \in \mathbb{R}^d$ under Sobolev smoothness constraints on the associated invariant density. Precisely, set

$$\Sigma_T(\beta, L; L', \sigma) := \Big\{\rho_b \mid b \in \Pi_0(\sigma),\ \rho_b \in \mathcal{S}(\beta+1, L'),\\ b^j \rho_b \in \mathcal{S}(\beta, L), 1 \leq j \leq d,\ \rho_b(x_0) \geq \rho_T^*\Big\},$$

where ρ_T^* is a sequence of positive real numbers such that $\lim_{T\to\infty} \rho_T^* = 0$ and $\liminf_{T\to\infty} \left(\rho_T^* \log T\right) > 0$. To shorten notation, we frequently write $\Sigma_T(\beta, L)$ for $\Sigma_T(\beta, L; L', \sigma)$. For constants $c_1 \in (0, \infty]$ and $c_2 > 0$, we further define $\Pi(c_1, c_2) = \Pi(c_1, c_2, \sigma)$ as the set of all drift functions $b \in \Pi_0(\sigma)$ satisfying the following conditions:

(P_1) It holds $\limsup_{\|x\|\to\infty} \|x\|^{-2} \langle b(x), x\rangle = -c_1$.

(P_2) For all $x \in \mathbb{R}^d$, we have $\|b(x)\| \leq c_2(1 + \|x\|)$.

A few comments on the definition of the functional sets $\Pi_0(\sigma)$ and $\Pi(c_1, c_2, \sigma)$ are in order:

Remark 1 • A lower bound on the value $\rho_b(x_0)$ is required for two reasons: First (and analogously to the case of nonparametric density estimation from i.i.d. observations considered in Butucea [3]), in order to obtain a reasonably good adaptive estimator of the value $(b^j \rho_b)(x_0)$, we have to exclude the case of a density ρ_b that varies with T such that $\rho_b(x_0) \to 0$ too fast. Secondly, for defining a ratio-type drift estimator in the spirit of (1.2), a strictly positive a priori lower bound $\rho_*(x_0) < \rho_b(x_0)$ is needed. The regularity conditions on the drift used in the proof of the asymptotic properties of our adaptive estimators actually allow for the derivation of explicit lower bounds; see Remark 2 below.

- The assumption of ergodicity is central for our subsequent analysis. Existence and uniqueness of invariant measures are conveniently proven by means of versions of Khasminskii's criterion, involving Lyapunov-type functions for the generator of the diffusion. Assumption (P_1) is a radial assumption on the drift coefficient and states that (if $c_1 < \infty$) the inward radial component of b has a prescribed polynomial behavior. In particular, (P_1) implies that $\exp(\delta\|x\|^2)$ for $\|x\| \geq 1$ is a Lyapunov function for small enough δ, thus ensuring the existence of an invariant measure. Together with the "at most linear growth"-condition in (P_2), it further implies an exponential bound on the associated invariant density (see Lemma 1 below).

3 Lower bound for pointwise estimation

In the Gaussian white noise framework, it has been shown by Lepski [12] that estimators which are optimally rate adaptive with respect to the pointwise risk over the scale of Hölder classes do not exist. The best adaptive estimators are proven to achieve only a rate which is slower than the optimal one in a logarithmic factor. Tsybakov [23] derives an analogous result for adaptation over the scale of Sobolev classes. To some extent, our findings are analogous, and principal ideas of the proof basically rely on techniques developed in the classical framework. The exact analysis of the drift estimation problem however also involves some subtleties which go beyond the known intricacies associated to the question of pointwise adaptation.

Let us first state the exact lower bound for estimating the components of $b\rho_b$ adaptively, assuming that the components $b^j\rho_b \in \mathcal{S}(\beta, L)$, $j = 1, \ldots, d$, for some $\beta \in [\beta_*, \infty)$ and $L \in [L_*, L^*]$. Here, $\beta_* \in (d/2, \infty)$ and $0 < L_* < L^* < \infty$ are fixed values. For any $\beta > d/2$, let

$$\kappa = \kappa(\beta) := \frac{\beta - \frac{d}{2}}{2\beta}, \quad \psi_{T,\beta} := \left(\frac{\log T}{T}\right)^{\kappa(\beta)}, \tag{3.1}$$

and recall the definition of $\mathbb{I}_\beta$ according to (1.4).

Theorem 1 *Fix $\beta_* > d/2$ and $\delta \in (0, 1)$, and denote $\mathcal{B}_T := [\beta_*, \beta_T] \times [L_*, L^*]$, for $\beta_T := (\log\log T)^\delta$. Then, for any $x_0 \in \mathbb{R}^d$ and $j \in \{1, \ldots, d\}$ fixed,*

$$\liminf_{T\to\infty} \inf_{\widehat{g}_T} \sup_{(\beta,L)\in\mathcal{B}_T} \sup_{b\in\Pi(c_1,c_2)} \sup_{\rho_b\in\Sigma_T(\beta,L;L',\sigma)} \frac{\mathbf{E}_b\left|\widehat{g}_T(x_0) - (b^j\rho_b)(x_0)\right|^2}{\psi_{T,\beta}^2 C_j^2(\beta, L; \rho_b, \sigma)} \geq 1, \tag{3.2}$$

where the infimum is taken over all estimators $\widehat{g}_T$ of $b^j\rho_b$ and

$$C_j(\beta, L; \rho_b, \sigma) := L^{\frac{d}{2\beta}} \frac{2\beta}{d} \left(\frac{d^2 a_{jj}\rho_b(x_0)}{\beta(2\beta - d)}\right)^{\frac{\beta-d/2}{2\beta}} \mathbb{I}_\beta. \tag{3.3}$$

The proof of Theorem 1 is deferred to Appendix B.1.

The basic—and classical—idea of the proof of Theorem 1 is to reduce the proof of the lower bound in (3.2) to proving a lower bound on the risk of two suitably chosen hypotheses. A lower bound on the latter risk is then deduced by means of Theorem 6(i) in Tsybakov [23] as it was also done in Butucea [3] and Klemelä and

Tsybakov [10,11]. The verification of the conditions of Tsybakov [23]'s result in the current diffusion framework however requires tools which differ from those used in the references mentioned above. Denoting by $\mathbb{P}_0$ and $\mathbb{P}_1$ the probability measures associated to the two different hypotheses, it needs to be shown that, for some fixed τ and for any $\alpha \in (0, 1/2)$,

$$\mathbb{P}_1\left(\frac{\mathrm{d}\mathbb{P}_0}{\mathrm{d}\mathbb{P}_1} \geq \tau\right) \geq 1 - \alpha. \tag{3.4}$$

In the Gaussian white noise framework considered in Klemelä and Tsybakov [10,11], (3.4) is verified directly for suitably chosen hypotheses due to the Gaussian nature of the model. For nonparametric density estimation from i.i.d. observations, Butucea [3] uses Lyapunov's CLT. In our framework, the condition (3.4) is verified by means of the martingale CLT.

4 Construction of sharp adaptive estimators

To define pointwise adaptive estimators of the components of $b\rho_b$ which attain the lower bound established in the previous section, we act similarly to Klemelä and Tsybakov [11]. Precisely, we will use a two-staged procedure in the spirit of Lepski's method, constructing first a collection of admissible estimators and selecting then an estimator with minimal variance among them. In contrast to the Gaussian white noise setting considered in Klemelä and Tsybakov [11], the complexity of the multidimensional diffusion model however requires a more involved investigation and more sophisticated tools. This remark applies both to the proof of asymptotic lower and upper bounds on the pointwise risk. In particular, for proving the exact upper bound, sufficiently precise exponential bounds on the stochastic error are needed. In the Gaussian white noise framework, the derivation of such exponential bounds is straightforward due to the Gaussian nature of the model. An additional complication arises in the classical problem of estimating a density at some fixed point $x_0 \in \mathbb{R}$ from i.i.d. observations (cf. Butucea [3]) where one has to derive exponential bounds on the risk which hold *uniformly* over a set of estimators associated to different bandwidths. To do so, Butucea [3] uses the classical Bernstein inequality and a uniform exponential inequality due to van de Geer [24]. Similarly to the pointwise density estimation problem, the bandwidths used for defining the estimators in our selection procedure involve an estimator $\widehat{\rho}_T(x_0)$ of the (unknown) value of the invariant density ρ_b at x_0 such that *uniform* risk bounds on the stochastic error are required.

We proceed by introducing central assumptions on the diffusion process X required for proving adaptivity of the proposed estimation scheme. Let P_t be the transition semigroup of X, and denote its transition density by p_t, i.e.

$$P_t f(x) = \mathbf{E}_b[f(X_t) \mid X_0 = x] = \int_{\mathbb{R}^d} f(y) p_t(x, y) \mathrm{d}y, \quad f \in L^2(\mu_b).$$

The following Bernstein-type deviation inequality in particular allows to prove *uniform* deviation inequalities which are crucial tools for verifying sufficiently sharp upper bounds on the pointwise squared risk of the adaptive estimators. Given any $b \in \Pi_0(\sigma)$, denote by $\varsigma_b^2(\cdot)$ the asymptotic variance appearing in the CLT, i.e.

$$\varsigma_b^2(g) := \lim_{T\to\infty} \frac{1}{T} \operatorname{Var}_{\mathbf{P}_b}\left(\int_0^T g(X_u)\mathrm{d}u\right), \quad g \in L^2(\mu_b). \tag{4.1}$$

Assumption (BI) Let $b \in \Pi_0(\sigma)$. Then there exists some positive constant C_B such that, for any bounded measurable function $f \in L^2(\mu_b)$ and for any $r, T > 0$ and $j \in \{1, \ldots, d\}$ fixed,

$$\mathbf{P}_b\left(\left|\frac{1}{T}\int_0^T f(X_u)b^j(X_u)\mathrm{d}u - \int_{\mathbb{R}^d} f(y)b^j(y)\mathrm{d}\mu_b(y)\right| > r\right) \le 2\exp\left(-\frac{Tr^2}{2C_B\left(\varsigma_b(f) + r\|f\|_\infty\right)}\right). \tag{BI}$$

The investigation of the variance term in (4.1) differs from the case of independent data as there appear additional covariances in the dependent case. The following assumption provides sufficiently tight upper bounds on the (asymptotic) variance.

Assumption (SG+) The carré du champs associated with the infinitesimal generator of the diffusion satisfies the spectral gap inequality, that is, for some constant c_P and any $f \in L^2(\mu_b)$,

$$\left\| P_t f - \int_{\mathbb{R}^d} f\mathrm{d}\mu_b \right\|_{L^2(\mu_b)} \le \mathrm{e}^{-t/c_P}\|f\|_{L^2(\mu_b)}. \tag{SG}$$

Furthermore, there exists some $C_0 > 0$ such that, for any $u \ge t > 0$ and for any pair of points $x, y \in \mathbb{R}^d$ with $\|x - y\|^2 \le u$, the transition density $p_t(\cdot, \cdot)$ satisfies

$$p_t(x, y) \ \le\ C_0\left(t^{-d/2} + u^{3d/2}\right). \tag{4.2}$$

For any symmetric diffusion X, it can be shown analogously to the proof of Proposition 1 in Dalalyan and Reiß [7] (also see the proof of Lemma 2.3 in Cattiaux et al. [4]) that, for any $f \in L^2(\mu_b)$ and $T > 0$,

$$\mathbf{E}_b\left[\left(\int_0^T f(X_u)\mathrm{d}u\right)^2\right] = 2\int_0^T\int_0^v \mathbf{E}_b\left[f(X_u)f(X_v)\right]\mathrm{d}u\ \mathrm{d}v \le 2T\int_0^T \langle f, P_w f\rangle_{L^2(\mu_b)}\,\mathrm{d}w.$$

The last term is upper-bounded by applying the Cauchy–Schwarz and the spectral gap inequality such that, for some positive constant C (depending only on c_P),

$$\mathbf{E}_b\left[\left(\int_0^T f(X_u)\mathrm{d}u\right)^2\right] \le CT\|f\|^2_{L^2(\mu_b)}. \tag{4.3}$$

It however turns out that, given the goal of describing the precise asymptotics for nonparametric drift estimation, we do actually require an exponential inequality with a tight leading term in the exponent. Taking also into account the upper bound on the transition density in (4.2), Proposition 1 in Dalalyan and Reiß [7] provides an enforced upper bound on the variance of additive functionals of multidimensional diffusions which allows to prove such a refined exponential inequality. In particular, for any compactly supported kernel $G : \mathbb{R}^d \to \mathbb{R}$, Assumption (**SG+**) ensures that there exists some positive constant C' (depending only on d, C_0 and c_P) such that, for any bandwidth $h > 0$, $y \in \mathbb{R}^d$, $T > 0$,

$$\operatorname{Var}_b\left(\frac{1}{\sqrt{T}}\int_0^T G_h(X_u - y)\mathrm{d}u\right) \leq C' \times \begin{cases} 1, & d = 1, \\ \max\left\{1, (\log(h^{-4}))^2\right\}, & d = 2, \\ h^{2-d}, & d \geq 3. \end{cases}$$

It seems to be rather unconventional to investigate estimators in diffusion models under the explicit assumption that functional inequalities in the spirit of the spectral gap hypothesis are satisfied. We believe that this approach is useful as it allows to formulate precise results for a sufficiently large class of diffusion processes under clear assumptions; see in particular Theorem 3 below.

The adaptive scheme is based on Lepski's principle. For implementing the procedure, consider a sufficiently fine grid $\mathcal{G} = \mathcal{G}_T$ on the interval $[\beta_*, \beta_T]$, with $\beta_T \to \infty$. It is defined as $\mathcal{G} = \mathcal{G}_T := \left\{\beta_1, \ldots, \beta_m\right\}$, where $\beta_* < \beta_1 < \cdots < \beta_m = \beta_T$. Assume that there exist $k_2 > k_1 > 0$ and $\delta_1 \geq \delta > 1$ such that

$$k_1(\log T)^{-\delta_1} \leq \beta_{i+1} - \beta_i \leq k_2(\log T)^{-\delta}, \quad i = 0, 1, \ldots, m-1, \tag{4.4}$$

and set $\beta_0 := \beta_* - d/2$. As in the case of density estimation from i.i.d. observations (cf. Butucea [3]), the optimal bandwidth for estimating $b^j\rho_b$ is not available in practice as it involves the unknown value of the invariant density ρ_b at $x_0 \in \mathbb{R}^d$. The adaptive procedure for estimating $b^j\rho_b$ therefore starts with a preliminary estimator $\widehat{\rho}_T(x_0)$ of the value $\rho_b(x_0)$.

Definition of the preliminary density estimator Define

$$\check{\rho}_T(x_0) := \frac{1}{Th_T^d}\int_0^T Q\left(\frac{X_u - x_0}{h_T}\right)\mathrm{d}u, \tag{4.5}$$

where Q is a bounded positive kernel satisfying $\int_{\mathbb{R}^d}\|u\||Q(u)|\mathrm{d}u < \infty$, and the bandwidth $h_T > 0$ is such that

$$\lim_{T\to\infty} h_T = 0, \quad \lim_{T\to\infty} Th_T^d = \infty, \quad \lim_{T\to\infty} Th_T^{2d}(\log T)^{-3} = \infty, \tag{4.6a}$$

and, for some $\alpha_0 \in (0, 1/2)$,

$$\limsup_{T\to\infty} h_T^d T^{\alpha_0} < \infty. \tag{4.6b}$$

Recall the definition of ρ_T^*, and let $\widehat{\rho}_T(x_0) := \max\left\{\check{\rho}_T(x_0),\ \rho_T^*\right\}$.

Main part of the procedure: adaptive estimation of $b\rho_b$ For fixed $j \in \{1, \ldots, d\}$, we now describe the procedure for defining an adaptive estimator of the j-th component of the vector $b\rho_b$. Recall that σ is the dispersion matrix taking values in $\mathbb{R}^{d\times d}$ and $a = \sigma\sigma^\top$ denotes the associated diffusion coefficient. The adaptive estimator will be selected among the family of estimators $\widehat{g}^j_{T,\beta}(x_0)$, defined as

$$\widehat{g}^j_{T,\beta}(x_0) := \frac{1}{T\widehat{h}^d_{T,\beta}} \int_0^T K_\beta\left(\frac{X_u - x_0}{\widehat{h}_{T,\beta}}\right) \mathrm{d}X^j_u,$$

where $\widehat{h}_{T,\beta} = \widehat{h}^j_{T,\beta} := \left(\frac{d\widehat{\rho}_T(x_0)a_{jj}\log T}{\beta T}\right)^{1/(2\beta)}$. As in Klemelä and Tsybakov [10], the kernel K_β is obtained as a renormalized version of the basic kernel

$$\widetilde{K}_\beta(x) := (2\pi)^{-d} \int_{\mathbb{R}^d} \left(1 + \|\lambda\|^{2\beta}\right)^{-1} \exp(\mathrm{i}\,\lambda^\top x)\mathrm{d}\lambda, \tag{4.7}$$

namely

$$K_\beta(x) := \mathrm{b}^d\,\widetilde{K}_\beta(\mathrm{b}x), \quad \text{for } \mathrm{b} = \mathrm{b}(\beta) := \left(\frac{2\beta - d}{d}\right)^{1/(2\beta)}. \tag{4.8}$$

As the last ingredient of the adaptive procedure, introduce the thresholding sequence

$$\widehat{\eta}_{T,\beta} = \widehat{\eta}^j_{T,\beta} := \left(\frac{d\widehat{\rho}_T(x_0)a_{jj}\log T}{\beta T}\right)^{\frac{\beta - d/2}{2\beta}} \|K_\beta\|_{L^2(\mathbb{R}^d)}.$$

The adaptive estimator $\widetilde{g}^j_T$ is defined as

$$\widetilde{g}^j_T(x_0) := \widehat{g}^j_{T,\widehat{\beta}^j_T}(x_0), \tag{4.9}$$

where

$$\widehat{\beta}^j_T := \max\left\{\beta \in \mathcal{G}_T :\ \left|\widehat{g}^j_{T,\gamma}(x_0) - \widehat{g}^j_{T,\beta}(x_0)\right| \le \widehat{\eta}_{T,\gamma}\ \forall\, \gamma \in \mathcal{G}_T,\ \gamma \le \beta\right\}. \tag{4.10}$$

We continue with the main result on pointwise adaptive estimation of $b\rho_b$. Recall the definition of the constants $C_j(\beta, L; \rho_b, \sigma)$, $j = 1, \ldots, d$, in (3.3). Denote by $\widetilde{\Pi}(c_1, c_2) = \widetilde{\Pi}(c_1, c_2, \sigma)$ the intersection of $\Pi(c_1, c_2, \sigma)$ with the set of all drift functions $b \in \Pi_0(\sigma)$ satisfying (BI) and Assumption (**SG+**).

Theorem 2 *For fixed* $\beta_* > d/2$, $0 < L_* < L^* < \infty$, $\delta \in (0, 1)$ *and for* $\mathcal{B}_T = [\beta_*, \beta_T] \times \left[L_*, L^*\right]$, *where* $\beta_T = (\log\log T)^\delta$, *the adaptive estimator* $\widetilde{g}^j_T$

defined according to (4.9) *satisfies, for any* $x_0 \in \mathbb{R}^d$,

$$\limsup_{T\to\infty} \sup_{(\beta,L)\in\mathcal{B}_T} \sup_{b\in\widetilde{\Pi}(c_1,c_2)} \sup_{\rho_b\in\Sigma_T(\beta,L;L',\sigma)} \frac{\mathbf{E}_b\left|\widetilde{g}_T^j(x_0)-(b^j\rho_b)(x_0)\right|^2}{\psi_{T,\beta}^2 C_j^2(\beta,L;\rho_b,\sigma)} \le 1.$$

The proof of Theorem 2 is given in Appendix B.2.

5 Sharp adaptive drift estimation for Kolmogorov diffusions

Restriction to the important case of Kolmogorov diffusions allows to derive and formulate results in a comparatively concise way.

Let $b \in \Pi_0(\sigma_0\,\mathbf{Id})$, $0 \neq \sigma_0 \in \mathbb{R}$, and consider the diffusion

$$X_t = X_0 + \int_0^t b(X_u)\mathrm{d}u + \sigma_0\, W_t, \qquad t \ge 0, \tag{5.1}$$

where W is a d-dimensional Brownian motion and the initial value X_0 is independent of W. Note that property (P_0') in the definition of the functional set Π_0 is fulfilled if $b = \sigma_0^2\nabla(\log\rho_b)/2$, that is, the drift vector b can be represented as a gradient. To enlighten notation, let

$$C(\beta,L;\rho_b,\sigma_0) = L^{\frac{d}{2\beta}}\frac{2\beta}{\sqrt{d}}\left(\frac{d^2\sigma_0^2\rho_b(x_0)}{\beta(2\beta-d)}\right)^{\frac{\beta-d/2}{2\beta}}\mathbb{I}_\beta. \tag{5.2}$$

We further introduce the maximal risk of an estimator $\check{g}_T$ of $b\rho_b$, for $\beta > d/2$, $L > 0$, $T > 0$, some bounded set $A \subset \mathbb{R}^d$ and fixed $x_0 \in \mathring{A}$ defined as

$$\mathscr{R}_{T,\beta,L}\left(\check{g}_T\right) := \sup_{b\in\widetilde{\Pi}(c_1,c_2)} \sup_{\rho_b\in\Sigma_T(\beta,L;L',\sigma_0\mathbf{Id})} \mathbf{E}_b\|\check{g}_T(x_0)-(b\rho_b)(x_0)\|^2. \tag{5.3}$$

Theorem 3 *Define* $\mathcal{B}_T$ *as in Theorem* 2, *and consider the risk introduced in* (5.3). *Then the following holds true:*

(a) *For any* x_0 *and for* $C(\beta,L;\rho_b,\sigma_0)$ *defined in* (5.2), *the estimator* $\widetilde{g}_T = \left(\widetilde{g}_T^j\right)_{j=1,\ldots,d}$ *defined according to* (4.9) *is sharp adaptive.*
(b) *If there exists an estimator* $\check{g}_T$ *such that, for some* $\beta_0 \ge \beta_*$, $L > 0$,

$$\limsup_{T\to\infty} \sup_{b\in\widetilde{\Pi}(c_1,c_2)} \sup_{\rho_b\in\Sigma_T(\beta_0,L;L',\sigma_0\mathbf{Id})} \frac{\mathbf{E}_b\|\check{g}_T(x_0)-(b\rho_b)(x_0)\|^2}{\psi_{T,\beta_0}^2 C^2(\beta_0,L;\rho_b,\sigma_0)} < 1,$$

then there exists $\beta_0' > \beta_0$ *such that*

$$\frac{\mathscr{R}_{T,\beta_0',L}\left(\check{g}_T\right)}{\mathscr{R}_{T,\beta_0',L}\left(\widetilde{g}_T\right)} \ge \Psi_T\,\frac{\mathscr{R}_{T,\beta_0,L}\left(\widetilde{g}_T\right)}{\mathscr{R}_{T,\beta_0,L}\left(\check{g}_T\right)}, \tag{5.4}$$

where $\Psi_T \to_{T\to\infty} \infty$. *In particular, for any fixed* $\beta \geq \beta_*$, $L > 0$,

$$\limsup_{T\to\infty} \sup_{b\in\widetilde{\Pi}(c_1,c_2)} \sup_{\rho_b\in\Sigma_T(\beta,L;L',\sigma_0\mathbf{Id})} \frac{\mathbf{E}_b\|\widetilde{g}_T(x_0)-(b\rho_b)(x_0)\|^2}{\psi_{T,\beta}^2 C^2(\beta,L;\rho_b,\sigma_0)} = 1.$$

The statement in the second part of the above theorem is to be interpreted in the sense that, whenever there exists an estimator $\check{g}_T$ which performs better than the estimator $\widetilde{g}_T$ at least for one smoothness degree β_0, there exists another smoothness factor β_0' for which there is much greater loss of $\check{g}_T$. The assertion—and its respective proof—are to be compared with Theorem 2 in Klemelä and Tsybakov [11].

Proof (of Theorem 3) We first show that $\Pi(c_1,c_2,\sigma_0\ \mathbf{Id}) = \widetilde{\Pi}(c_1,c_2,\sigma_0\ \mathbf{Id})$. Let $b \in \Pi(c_1,c_2,\sigma_0\ \mathbf{Id})$. In view of the results in Section 4.3 in Bakry et al. [2] (p. 747), (P_1) implies that (SG) holds. Since, in addition, (P_2) is satisfied, Theorem 3.2 in Qian and Zheng [17] entails that (4.2) and thus Assumption (**SG+**) is fulfilled. For any $b \in \Pi_0(\sigma_0\ \mathbf{Id})$, the associated measure μ_b is reversible for X (see, e.g., Lemma 2.2.3 in Royer [19]). In particular, (SG) is equivalent to Poincaré's inequality. Restricting to bounded drift functions, Poincaré's inequality implies that Assumption (BI) is satisfied, too; this follows from Lemma 2 stated in Appendix A. In view of Theorem 1 and Theorem 2, it now only remains to verify (5.4). The proof actually is along the lines of the proof of Theorem 2 in Klemelä and Tsybakov [11] and therefore omitted. □

We conclude this section with a brief summary of the adaptive estimation procedure. Assume that a continuous record of observations $X^T = (X_t)_{0\leq t\leq T}$ of a diffusion process solution of the SDE (2.1) is available and that the (constant) diffusion matrix $a = \sigma\sigma^\top$ is known. The goal is to estimate the value of the j-th component of the product $b\rho_b$ at some given fixed point x_0. For implementing the adaptive estimation scheme, we need to specify a lower bound $\beta_* > d/2$ on the unknown smoothness of the function $b\rho_b$ and fix some value $\delta \in (0,1)$. To define an estimator on the basis of the input parameters X^T, $a = (a_{ij})_{1\leq i,j\leq d}$, x_0, β_* and δ, one then might proceed as follows:

∗ (*Computation of a pilot estimator of the invariant density*) Choose some bounded positive kernel $Q\colon \mathbb{R}^d \to \mathbb{R}$ satisfying $\int_{\mathbb{R}^d}\|u\||Q(u)|\mathrm{d}u < \infty$ and some bandwidth $h_T > 0$ fulfilling (4.6). Define a preliminary density estimator $\widehat{\rho}_T(x_0)$ by computing $\check{\rho}_T(x_0)$ according to (4.5) and by letting

$$\widehat{\rho}_T(x_0) := \max\left\{\check{\rho}_T(x_0),\ \rho_T^*\right\},$$

where ρ_T^* denotes a vanishing sequence of positive real numbers satisfying $\liminf_{T\to\infty}(\rho_T^*\log T) > 0$.

∗ (*Computation of kernel estimators for a discrete set of parameters*) Specify a grid $\mathcal{G}_T := \{\beta_1,\ldots,\beta_m\}$, where the values $\beta_* < \beta_1 < \cdots < \beta_m = (\log\log T)^\delta$ are chosen such that (4.4) is satisfied. Define the bandwidths

$$\widehat{h}_{T,\beta_i} = \left(\frac{d\widehat{\rho}_T(x_0)a_{jj}\log T}{\beta_i T}\right)^{\frac{1}{2\beta_i}}, \quad i = 1,\ldots,m.$$

Recall the definition of the kernel K_β in (4.8), and compute the family of estimators

$$\widehat{g}^j_{T,\beta_i}(x_0) = \frac{1}{T\widehat{h}^d_{T,\beta_i}} \int_0^T K_{\beta_i}\left(\frac{X_u - x_0}{\widehat{h}_{T,\beta_i}}\right) \mathrm{d}X^j_u, \quad i = 1, \ldots, m. \tag{5.5}$$

∗ (*Definition of the Lepski-type estimator of* β *and the adaptive estimator of* $b^j\rho_b$) Recall the definition of $\mathbb{I}^2_\beta$ in (1.4), and define the thresholding values

$$\widehat{\eta}_{T,\beta_i} = \left(\frac{d\widehat{\rho}_T(x_0)a_{jj}\log T}{\beta_i T}\right)^{\frac{\beta_i - d/2}{2\beta_i}} \mathbb{I}_{\beta_i}\left(\frac{2\beta_i - d}{d}\right)^{\frac{\beta_i + d/2}{2\beta_i}}, \quad i = 1, \ldots, m. \tag{5.6}$$

Use the values (5.5) and (5.6) to determine $\widehat{\beta}^j_T$ as specified in (4.10), and define the adaptive estimator $\widetilde{g}^j_T(x_0) = \widehat{g}^j_{T,\widehat{\beta}^j_T}(x_0)$.

Remark 2
- Once one has determined the adaptive estimator $\widetilde{g}_T = (\widetilde{g}^j_T)_{j=1,\ldots,d}$ according to the above scheme, one obtains an adaptive drift estimator by defining a suitable invariant density estimator $\widetilde{\rho}_T$ and setting

$$\widetilde{b}_T(x_0) := \frac{\widetilde{g}_T(x_0)}{\widetilde{\rho}_T(x_0) \vee \rho_*(x_0)}, \qquad x_0 \in \mathbb{R}^d,$$

  $\rho_*(x_0) > 0$ denoting some a priori lower bound on $\rho_b(x_0)$. Restricting again to the case of Kolmogorov diffusions as in (5.1), the normalizing factor appearing in the upper bound for the pointwise squared risk of $\widetilde{b}^j_T(x_0)$, assuming that $b \in \Pi(c_1, c_2)$ and $\rho_b \in \Sigma_T(\beta, L; L', \sigma_0\ \mathbf{Id})$, is identified as $C_j(\beta, L; \rho_b, \sigma_0)\rho_b^{-1}(x_0) = D_j(\beta, L; \rho_b, \sigma_0)$.
- In the situation of Theorem 3, an explicit a priori lower bound on $\rho_b(x_0)$ depending only on c_1, c_2 can be derived as in Remark 6 in Dalalyan and Reiß [7]. For the more general case of diffusion processes with uniformly nondegenerate diffusion matrix a, it was proven in Metafune et al. [15] that, if $b^i \in C^2(\mathbb{R}^d)$, ($P_1$) is satisfied, and, in addition, $\|b(x)\| + \|Db(x)\| + \|D^2b(x)\| \leq c_1'(1 + \|x\|)$ for some constant $c_1' > 0$, it holds $\rho_b(x) \geq \mathrm{e}^{-K(1+\|x\|^2)}$, $x \in \mathbb{R}^d$, with some positive constant K depending only on c_1', c_2, d and σ.

6 Discussion and extensions

Placement and interpretation of the identified constants Let us first arrange the normalizing factors identified in the previous sections and relate it to known results on asymptotically exact adaptive estimation with respect to pointwise risk over the scale of Sobolev classes in the classical statistical models. Tsybakov [23] considers the problem of nonparametric function estimation in the Gaussian white noise model (in the one-dimensional case), assuming that the unknown function belongs to some Sobolev class with unknown regularity parameter. The question of density estimation at a fixed point $x_0 \in \mathbb{R}$ is investigated by Butucea [3]. Since the variance of the proposed kernel estimator is proportional to the value of the unknown density f at x_0, the value $f(x_0)$ appears in the exact normalization. Klemelä and Tsybakov [11] deal with nonparametric estimation of a multivariate function and its partial derivatives at a fixed point when the Riesz transform is observed in Gaussian white noise. In particular, Klemelä and Tsybakov [11] find the exact constant for nonparametric estimation of a function $f : \mathbb{R}^d \to \mathbb{R}$, observed in Gaussian white noise and satisfying the Sobolev smoothness constraint $\eta_\beta(f) \leq L$. In combination with the results of Butucea [3] on classical density estimation (in dimension $d = 1$), the exact constant for nonparametric estimation of a density $f : \mathbb{R}^d \to \mathbb{R}$ at some fixed point $x_0 \in \mathbb{R}^d$ is then identified as

$$L^{\frac{d}{2\beta}} \frac{2\beta}{d} \left(\frac{d^2 f(x_0)}{\beta(2\beta - d)} \right)^{\frac{\beta - d/2}{2\beta}} \mathbb{I}_\beta. \tag{6.1}$$

For the case of Kolmogorov diffusions with $\sigma\sigma^\top = \mathbf{Id}$, $C_j(\beta, L; \rho_b, \mathbf{Id})$ as introduced in (3.3) coincides with the constant in (6.1). The accordance of constants in the exact asymptotics reflects the old-established experience that different statistical models show similar behavior, at least from an asymptotic point of view. The remarkable results of Dalalyan and Reiß [7] on asymptotic statistical equivalence for inference on the drift in multidimensional Kolmogorov diffusion models justify this notice in a mathematically formal manner. They are however established under rather restrictive (Hölder) smoothness assumptions. Precisely, the critical regularity for proving asymptotic equivalence with the Gaussian shift model grows like $(1/2 + 1/\sqrt{2})d$ as $d \to \infty$. The authors refer to the question whether for Hölder classes of smaller regularity asymptotic equivalence fails as "a challenging open problem." Our risk bounds are valid under weaker smoothness constraints which suggests that asymptotic equivalence (at least in a reduced sense) still holds beyond the critical bounds of Dalalyan and Reiß [7]. Going beyond the special case of Kolmogorov diffusions with unit diffusion part, the dependence of the drift estimation problem on the form of the diffusion coefficient $a = \sigma\sigma^\top$ is reflected by the appearance of the j-th diagonal entry of the matrix a in the optimal normalizing factor $C_j(\beta, L; \rho_b, \sigma)$ (for estimating $b^j\rho_b$).

Possible generalizations of the assumptions It can be argued which type of description of the dependence structure underlying the diffusion is most convenient. Given the goal of describing exact asymptotics for pointwise drift estimation for an as large as possible class of diffusion processes under some preferably small set of assumptions, we decided to formulate our results in terms of the spectral gap hypothesis. Indeed, restricting to the case of reversible diffusion processes, Theorem 3.1 in Guillin et al. [8] implies that, whenever $\varsigma_b^2(g) \leq C\|g\|_\infty^2$ for any centered bounded g and some constant $C > 0$, (BI) entails the Poincaré inequality, i.e.

$$\mathrm{Var}_{\mu_b}(f) = \int f^2 \mathrm{d}\mu_b - \left(\int f \mathrm{d}\mu_b \right)^2 \leq c_P^{-1} \int |\nabla f|^2 \mathrm{d}\mu_b \tag{PI}$$

for any smooth enough function $f : \mathbb{R}^d \to \mathbb{R}$ and some positive constant c_P. It is further well-known that Poincaré's inequality in the symmetric case is equivalent to the spectral gap assumption (SG). However, as was already noted in the introduction, the upper bound result in Theorem 2 essentially holds whenever a Bernstein-type deviation inequality in the spirit of (BI) is combined with a sufficiently tight upper variance bound [as provided by Assumption (**SG+**)]. Such variance bounds for diffusion processes actually can be deduced by means of mixing conditions or the assumption of weak dependence of data.

Acknowledgments The author is grateful to her Ph.D. advisor, Angelika Rohde, and Enno Mammen for constant encouragement and constructive advise. She also would like to thank two anonymous referees for helpful comments which led to a substantial improvement of the paper. This work was partially supported by the DFG Priority Program SPP 1324 (Project: RO 3766/2-1).

Appendix A: Preliminaries

We first give a result which allows to deduce the exact asymptotics for pointwise estimation of the drift component b^j from exact results on estimating $b^j\rho_b$, $j \in \{1, \dots, d\}$. In particular, it allows to identify $D(\beta, L; \rho_b, \sigma_0)$ as defined in (1.3) as the optimal normalizing factor for estimating the j-th component of $b \in \Pi(c_1, c_2, \sigma_0\mathbf{Id})$. For the detailed derivation of the exact lower bound, we refer to Theorem 2.5.7 in Strauch [21].

Lemma 1 (a) *There exist two positive constants C_1, C_2 (depending only on c_1, c_2 and σ) such that the invariant density ρ_b satisfies $\rho_b(x) \le C_1 \mathrm{e}^{-C_2\|x\|^2}$, $x \in \mathbb{R}^d$, for any $b \in \Pi(c_1, c_2)$.*

(b) *If $b \in \widetilde{\Pi}(c_1, c_2)$ and if $\rho_b \in \mathcal{S}(\beta + 1, L')$, for some $\beta > d/2$ and $L' > 0$, then there exists an invariant density estimator $\widehat{\rho}_T$ such that*

$$\mathbf{E}_b|\widehat{\rho}_T(x) - \rho_b(x)|^2 \le K_1 T^{-\frac{\beta+1-(d/2)}{\beta+1}} \exp(-K_2\|x\|), \qquad x \in \mathbb{R}^d, \qquad (7.1)$$

where the constants K_1, K_2 depend only on L', c_1, c_2 and σ.

Proof (a) The pointwise upper bound on ρ_b is an immediate consequence of the results of Metafune et al. [15] who study global regularity properties of invariant measures of divergence-form operators. Their results also hold in our specific framework since we restrict attention to the case of constant, uniformly elliptic diffusion part. Denote by $\lambda_{\max}$ the largest eigenvalue of a. Due to Corollary 2.5 in Metafune et al. [15], (P_1) implies that $\exp(\eta\|x\|^2) \in L^1(\mu_b)$ for $\eta < c_1\,(2\lambda_{\max})^{-1}$. Since $\|b(x)\| \le c_2(1 + \|x\|) \lesssim \exp(\|x\|)$, Theorem 6.1 in Metafune et al. [15] applies and yields the assertion.

(b) Let $G = G_{\beta+1} : \mathbb{R}^d \to \mathbb{R}$ be the kernel with Fourier transform

$$\phi_G(\lambda) = \int_{\mathbb{R}^d} \mathrm{e}^{\mathrm{i}\lambda^\top y} G(y)\mathrm{d}y = \frac{1}{1 + \|\lambda\|^{2(\beta+1)}}, \quad \lambda \in \mathbb{R}^d,$$

and define the invariant density estimator

$$\widehat{\rho}_T(x) = \widehat{\rho}_{T,h}(x) := \frac{1}{Th^d}\int_0^T G\left(\frac{X_u - x}{h}\right)\mathrm{d}u, \quad x \in \mathbb{R}^d.$$

The bandwidth $h = h_T \searrow 0$ is to be specified later. For bounding the stochastic error, note that, using (4.3),

$$\begin{aligned}\mathbf{E}_b\left|\widehat{\rho}_T(x) - \mathbf{E}_b\widehat{\rho}_T(x)\right|^2 &= \frac{1}{T^2h^{2d}}\ \mathrm{Var}_b\left(\int_0^T G\left(\frac{X_u - x}{h}\right)\mathrm{d}u\right)\\ &\leq \frac{C}{Th^{2d}}\int_{\mathbb{R}^d} G^2\left(\frac{y-x}{h}\right)\rho_b(y)\mathrm{d}y.\end{aligned}$$

Taking into account the regularity properties of G, a multidimensional version of Theorem 1A in Parzen [16] yields

$$\mathbf{E}_b\left|\widehat{\rho}_T(x) - \mathbf{E}_b\widehat{\rho}_T(x)\right|^2 \leq \frac{C}{Th^d}\rho_b(x)\|G\|^2_{L^2(\mathbb{R}^d)}(1 + o_T(1)).$$

It remains to treat the bias term. Note that, using in particular Cauchy–Schwarz,

$$\begin{aligned}\left|\mathbf{E}_b\widehat{\rho}_T(x) - \rho_b(x)\right| &= (2\pi)^{-d}\left|\int_{\mathbb{R}^d}\phi_{\rho_b}(\lambda)\left\{\left(1 + \|h\lambda\|^{2\beta}\right)^{-1} - 1\right\}\mathrm{e}^{-\mathrm{i}\lambda^\top x}\mathrm{d}\lambda\right|\\ &\leq h^{\beta+1}\left((2\pi)^{-d}\int_{\mathbb{R}^d}\left|\phi_{\rho_b}(\lambda)\right|^2\|\lambda\|^{2(\beta+1)}\mathrm{d}\lambda\right)^{1/2}\\ &\quad\times\left((2\pi)^{-d}\int_{\mathbb{R}^d}\frac{\|h\lambda\|^{2(\beta+1)}}{\left(1 + \|h\lambda\|^{2(\beta+1)}\right)^2}\mathrm{d}\lambda\right)^{1/2}\\ &\leq L'\,(2\pi)^{-d/2}\left(\int_{\|y\|\leq 1}\frac{\mathrm{d}y}{(1 + \|y\|^{2\beta})^2} + \int_{\|y\|>1}\frac{\mathrm{d}y}{\|y\|^{2\beta}}\right)^{1/2}\\ &\quad\times h^{\beta+1-d/2} =: Mh^{\beta+1-d/2}.\end{aligned}$$

Specifying $h = h_T \sim \left(\frac{C\rho_b(x)}{M^2T}\right)^{\frac{1}{2(\beta+1)}}$ and using the upper bound on $\rho_b(x)$ from part (a), we obtain (7.1). □

Denote by $N_{[\,]}\left(\varepsilon, \mathcal{F}, L^2(\mu_b)\right)$ the ε-entropy with bracketing, that is, the smallest number of ε-brackets (in $L^2(\mu_b)$) which are required to cover $\mathcal{F}$ (cf. van der Vaart and Wellner [25], Definition 2.1.6).

Lemma 2 (a) *Let $b \in \Pi(c_1, c_2, \sigma)$, and suppose that X satisfies (PI). Fix $j \in \{1, \ldots, d\}$, and assume that there exists some positive constant B such that, for any bounded measurable $f \in L^2(\mu_b)$,*

$$\max\left\{\sup_{x\in\mathrm{supp}(f)}|b^j(x)|, \sup_{x\in\mathrm{supp}(f^2)}|b^j(x)|^2\right\} \leq B. \tag{7.2}$$

Then Assumption (BI) is satisfied.

(b) *Let $\mathcal{F} \subset L^2(\mu_b)$ be some class of measurable functions $f : \mathbb{R}^d \to \mathbb{R}$, and assume that, for some positive constants K and M, it holds*

$$\sup_{f \in \mathcal{F}} \|f\|_\infty \le K, \quad \sup_{f \in \mathcal{F}} \|f\|_{L^2(\mu_b)} \le M.$$

Grant Assumptions (BI) and (SG). Then, for arbitrary $T > 0$ and any positive r satisfying, for some positive constants K_1 and K_2,

$$\frac{K_1}{\sqrt{T}} \int_0^1 \max\left\{\sqrt{\log N_{[\,]}(\varepsilon, \mathcal{F}, L^2(\mu_b))}, 1\right\} \mathrm{d}\varepsilon \;\le\; r \;\le\; \frac{K_2 M^2}{K},$$

there exist some positive constants C_1 and C_2 such that

$$\mathbf{P}_b\left(\sup_{f \in \mathcal{F}} \left|\frac{1}{T}\int_0^T f(X_u)\mathrm{d}X_u^j - \int_{\mathbb{R}^d} f(y)b^j(y)\mathrm{d}\mu_b(y)\right| > r\right) \le C_1 \exp\left(-\frac{C_2 T r^2}{M^2}\right). \tag{BI+}$$

Proof (a) Letting, for $r, T > 0$,

$$\mathsf{p}_T(r) := \mathbf{P}_b\left(\left|\frac{1}{T}\int_0^T \left(f(X_u)b^j(X_u) - \int_{\mathbb{R}^d} f(y)b^j(y)\mathrm{d}\mu_b(y)\right)\mathrm{d}u\right| > r\right), \tag{7.3}$$

Theorem 1.1 in Lezaud [13] implies that

$$\mathsf{p}_T(r) \le 2\exp\left(-\frac{Tr^2}{2\left(\varsigma_b^2(fb^j) + c_P\|fb^j\|_\infty r\right)}\right). \tag{7.4}$$

Using the spectral gap assumption, we get, for any $T > 0$, $g \in L^2(\mu_b)$,

$$\frac{1}{T}\mathrm{Var}_{\mathbf{P}_b}\left(\int_0^T g(X_u)\mathrm{d}u\right) \le 2\int_0^T \langle P_t g, g\rangle_{\mu_b}\mathrm{d}t \le 2\|g\|^2_{L^2(\mu_b)} \int_0^T \mathrm{e}^{-2t/c_P}\mathrm{d}t \le c_P\|g\|^2_{L^2(\mu_b)}.$$

Consequently, in view of (7.2),

$$\varsigma_b^2(fb^j) = \lim_{T\to\infty}\frac{1}{T}\mathrm{Var}_{\mathbf{P}_b}\left(\int_0^T (fb^j)(X_u)\mathrm{d}u\right) \le c_P\|fb^j\|^2_{L^2(\mu_b)} \le c_P B\|f\|^2_{L^2(\mu_b)}$$

and $\|fb^j\|_\infty \le B\|f\|_\infty$. Plugging these estimates into (7.4), we obtain the asserted inequality.

(b) Under the given assumptions, Bernstein's inequality for continuous martingales can be used to show that there exists some constant $\widetilde{C}_B$ such that, for any $r, T > 0$,

$$\begin{aligned}\mathsf{q}_T(r) &:= \mathbf{P}_b\left(\left|\frac{1}{T}\int_0^T f(X_u)\mathrm{d}X_u^j - \int_{\mathbb{R}^d} f(y)b^j(y)\mathrm{d}\mu_b(y)\right| > r\right)\\ &\leq 2\exp\left(-\frac{Tr^2}{2\widetilde{C}_B\left(\|f\|^2_{L^2(\mu_b)} + \|f\|_\infty\right)}\right). \end{aligned} \tag{7.5}$$

To see this, write $\mathsf{q}_T(r) \leq \mathsf{p}_T(r/2) + \mathsf{p}_T'(r/2)$, for $\mathsf{p}_T(\cdot)$ introduced in (7.3) and

$$\mathsf{p}_T'(r) := \mathbf{P}_b\left(\left|\frac{1}{T}\int_0^T f(X_u)\sum_{k=1}^d \sigma_{jk}\mathrm{d}W_u^k\right| > r\right), \quad r > 0.$$

Letting $M_t(f) := \int_0^t f(X_u)\sum_{k=1}^d \sigma_{jk}\ \mathrm{d}W_u^k$, $t \geq 0$, and denoting by $\langle M\rangle_\cdot$ the quadratic variation of the martingale M, Bernstein's inequality for continuous martingales (see p. 154 in Revuz and Yor [18]) gives

$$\begin{aligned}\mathsf{p}_T'(r/2) &\leq \mathbf{P}_b\Big(\big|M_T(f)\big| > Tr/2;\ \langle M(f)\rangle_T \leq Tr\|f\|_\infty/2\Big)\\ &\quad + \mathbf{P}_b\Big(\langle M(f)\rangle_T > Tr\|f\|_\infty/2\Big)\\ &\leq 2\exp\left(-\frac{Tr}{4\|f\|_\infty}\right) + \underbrace{\mathbf{P}_b\left(T^{-1}\int_0^T f^2(X_u)\mathrm{d}u > a_{jj}^{-1}r\|f\|_\infty/2\right)}_{=:\mathsf{p}_T''(r)}.\end{aligned}$$

Theorem 1.1 in Lezaud [13] then can be used to show that

$$\mathsf{p}_T''(r) \leq \exp\left(-\frac{Tr^2}{8c_P a_{jj}\left(a_{jj}\ \|f\|^2_{L^2(\mu_b)} + \|f\|_\infty/2\right)}\right).$$

The inequality (7.5) now follows for $\widetilde{C}_B := 4\max\left\{2c_P, 2c_P a_{jj}^2, c_P a_{jj}, 1\right\}$. In view of (7.5), a uniform exponential inequality in the spirit of Theorem 5.11 in van de Geer [24] is available. Indeed, Theorem 5.11 in van de Geer [24] appears as a special case of the uniform inequality for martingales in van de Geer [24]'s Theorem 8.13, and the proof of Theorem 8.13 continues to hold in the diffusion setting if the Bernstein inequality for martingales in van de Geer [24]'s Corollary 8.10 is replaced with the Bernstein-type deviation inequality (BI). □

For the proof of the following Lemma, we refer to the proofs of Proposition 1 in Klemelä and Tsybakov [10] and of Lemma 10 in Tsybakov [23].

Lemma 3 *Let $\beta > d/2$, and, for $\mathbb{I}_\beta$, $\widetilde{K}_\beta(\cdot)$ and $\mathrm{b} = \mathrm{b}(\beta)$ defined according to* (1.4)*,* (4.7) *and* (4.8)*, respectively, let $K_\beta^*(x) := \mathbb{I}_\beta^{-1}\mathrm{b}^{-\beta+d/2}\ \widetilde{K}_\beta(\mathrm{b}x)$.*

(a) (cf. Proposition 1 in *Klemelä and Tsybakov* [10]) *It holds*

$$\widetilde{K}_\beta(0) = (2\pi)^{-d} \int_{\mathbb{R}^d} \left(1 + \|\lambda\|^{2\beta}\right)^{-1} \mathrm{d}\lambda = \frac{2\beta}{d}\, \mathbb{I}_\beta^2,$$

and, for $K_\beta(x) = \mathrm{b}^d \widetilde{K}_\beta(\mathrm{b}x)$, $\left\|K_\beta\right\|_{L^2(\mathbb{R}^d)} = \mathbb{I}_\beta \left(\frac{2\beta-d}{d}\right)^{\frac{\beta+d/2}{2\beta}} = \mathbb{I}_\beta\, \mathrm{b}^{\beta+d/2}$.

(b) (cf. Lemma 10 in *Tsybakov* [23]) *For fixed* $\delta \in (0, 1)$, *there exists some compactly supported modification* $\overline{K}_\beta$ *of* K_β^* *which enjoys the following properties,*

$$\left\|\overline{K}_\beta\right\|_{L^2(\mathbb{R}^d)} \leq 1 - \delta/2, \tag{K1}$$

$$\eta_\beta(\overline{K}_\beta) \leq 1 - \delta/2, \tag{K2}$$

$$(1 - \delta/2) K_\beta^*(0) \leq \overline{K}_\beta(0) \leq K_\beta^*(0). \tag{K3}$$

Appendix B: Proofs

B.1: Lower bound

Proof (of Theorem 1) Let $\psi_{\beta,L} = \psi_{\beta,L}^j := \psi_{T,\beta} C_j(\beta, L; \rho_b, \sigma)$, for $\psi_{T,\beta}$ and $C_j(\beta, L; \rho_b, \sigma)$ defined in (3.1) and (3.3), respectively. To enlighten notation, the dependence on the coordinate $j \in \{1, \ldots, d\}$ will be mostly suppressed in the sequel.

(I) Construction of the hypotheses. Let $L \in [L_*, L^*]$, fix some nondegenerate $\mathbb{R}^{d\times d}$-matrix σ, let $a := \sigma\sigma^\top$, and consider some positive density function

$$\rho \in C^\infty(\mathbb{R}^d) \cap \mathcal{S}\left(\beta_T + 1, L'\right) \cap \mathcal{S}\left(\beta_* + 1, L'\right), \quad \text{where } L' := 2L \left(\sum_{k=1}^d a_{jk}^2\right)^{-1/2}.$$

Fix $\delta_0 \in (0, 1/2)$, $c_1, c_2 > 0$, and assume that ρ is such that the function

$$\rho_{T,0}(x) := \delta_0^{1/(\beta_*+3/2)}\, \rho\left(x \delta_0^{1/(\beta_*+3/2)}\right), \quad x \in \mathbb{R}^d,$$

satisfies $\rho_{T,0}(x_0) \geq \rho_T^*$, for any x with $\|x\|$ large enough,

$$\langle a \nabla \log \rho_{T,0}(x),\ x\rangle \leq -2c_1 \|x\|^2,$$

and for any $x \in \mathbb{R}^d$,

$$\left|\sum_{k=1}^d a_{jk} \partial_k \rho_{T,0}(x)\right| \leq 2 \quad \text{and} \quad \|\nabla \log \rho_{T,0}(x)\| \leq 2\|a\|_{S_2}^{-1} c_2.$$

Consequently, $a\nabla(\log \rho_{T,0})/2 =: b_{T,0} \in \Pi(c_1, c_2, \sigma)$. In particular, the SDE $\mathrm{d}X_t = b_{T,0}(X_t)\mathrm{d}t + \sigma\ \mathrm{d}W_t,\ t \geq 0$, admits a strong solution with Lebesgue continuous invariant measure and invariant density $\rho_{T,0}$. Define further

$$g_{T,0}(x) := \frac{1}{2}\sum_{k=1}^{d} a_{jk}\partial_k \rho_{T,0}(x), \quad x \in \mathbb{R}^d.$$

For $\beta > d/2$, consider $\widetilde{K}_\beta$ and $\mathrm{b} = \mathrm{b}(\beta)$ as introduced in (4.7) and (4.8), respectively, and denote again $K_\beta^*(x) = \mathbb{I}_\beta^{-1}\mathrm{b}^{-\beta+d/2}\ \widetilde{K}_\beta(\mathrm{b}x)$. Lemma 3 implies that

$$K_\beta^*(0) = \mathbb{I}_\beta^{-1}\mathrm{b}^{-\beta+d/2}\ \widetilde{K}_\beta(0) = \frac{2\beta}{d}\left(\frac{d}{2\beta - d}\right)^{\frac{\beta-d/2}{2\beta}}\ \mathbb{I}_\beta$$

and

$$\|K_\beta^*\|_{L^2(\mathbb{R}^d)} = \mathbb{I}_\beta^{-1}\mathrm{b}^{-\beta+d/2}\left(\int_{\mathbb{R}^d}\widetilde{K}_\beta^2(\mathrm{b}x)\mathrm{d}x\right)^{1/2} = 1.$$

Denote by $\overline{K}_\beta$ the compactly supported modification of K_β^* from Lemma 3 satisfying (K1), (K2), and (K3) for $\delta = \delta_0$. Define the function $g_{T,\beta_*} : \mathbb{R}^d \to \mathbb{R}$ such that, for any $k \in \{1, \ldots, d\}$,

$$\partial_k g_{T,\beta_*}(x) = 2La_{jj}^{-1}h_{T,\beta_*}^{\beta_*-d/2}\ \overline{K}_{\beta_*}\left(\frac{x - x_0}{h_{T,\beta_*}}\right)\ \delta_{kj}, \qquad x \in \mathbb{R}^d, \tag{8.1}$$

where

$$h_{T,\beta_*} := \left(\frac{d\rho_{T,0}(x_0)a_{jj}\log T}{\beta_* L^2 T}\right)^{1/(2\beta_*)}. \tag{8.2}$$

Let

$$\rho_{T,1}(x) := \rho_{T,0}(x)\left(1 - \int_{\mathbb{R}^d} g_{T,\beta_*}(y)\mathrm{d}y\right) + g_{T,\beta_*}(x),$$

and consider the hypothesis $g_{T,1}$, defined as $g_{T,1} := \sum_{k=1}^d a_{jk}\partial_k\rho_{T,1}/2$. The function $b_{T,1} : \mathbb{R}^d \to \mathbb{R}^d$ is taken as $b_{T,1} := a\nabla(\log \rho_{T,1})/2$. Note that, for T large enough,

$$\frac{\rho_{T,0}(x_0)}{\rho_{T,1}(x_0)} \leq \frac{\rho_{T,0}(x_0)}{\rho_{T,0}(x_0)\left(1 - \int_{\mathbb{R}^d} g_{T,\beta_*}(y)\mathrm{d}y\right)} \leq 1 + \delta_0/2. \tag{8.3}$$

Plugging in the respective definitions of the hypotheses, it can be shown that $g_{T,0} \in \mathcal{S}(\beta_T, L)$, $g_{T,1} \in \mathcal{S}(\beta_*, L)$ and $b_{T,1} \in \Pi(c_1, c_2, \sigma)$. The above definitions of the hypotheses further imply that $\rho_{T,0} \in \mathcal{S}(\beta_T + 1, L')$, $\rho_{T,1} \in \mathcal{S}(\beta_* + 1, L')$ and

$$2b^j_{T,i}\rho_{T,i} = \sum_{k=1}^d a_{jk}\partial_k\big(\log\rho_{T,i}\big)\rho_{T,i} = \sum_{k=1}^d a_{jk}\partial_k\rho_{T,i}, \quad i\in\{0,1\}.$$

Summing up, $\rho_{T,0}\in\Sigma_T(\beta_T,L)$ and $\rho_{T,1}\in\Sigma_T(\beta_*,L)$.

(II) A VERSION OF THEOREM 6(i) IN Tsybakov [23]. The central ingredient of the proof is a special case of Theorem 6(i) in Tsybakov [23]. It will be applied in the following situation: Denote by $\mathbf{E}_i = \mathbf{E}_{b_{T,i}}$ expectation under the measure $\mathbf{P}_i = \mathbf{P}_{b_{T,i}}$ associated with the hypothesis $b = b_{T,i}$, $i\in\{0,1\}$, and note that

$$\begin{aligned}
&\inf_{\widehat{g}_T}\sup_{(\beta,L)\in\mathcal{B}_T}\sup_{b\in\Pi(c_1,c_2,\sigma)}\sup_{\rho_b\in\Sigma_T(\beta,L)}\psi^{-2}_{\beta,L}\,\mathbf{E}_b\big|\widehat{g}_T(x_0)-(b^j\rho_b)(x_0)\big|^2\\
&\quad\geq \inf_{\widehat{g}_T}\max\Bigg\{\sup_{b\in\Pi(c_1,c_2,\sigma)}\sup_{\rho_b\in\Sigma_T(\beta_T,L)}\psi^{-2}_{\beta_T,L}\,\mathbf{E}_b\big|\widehat{g}_T(x_0)-(b^j\rho_b)(x_0)\big|^2,\\
&\qquad\qquad\sup_{b\in\Pi(c_1,c_2,\sigma)}\sup_{\rho_b\in\Sigma_T(\beta_*,L)}\psi^{-2}_{\beta_*,L}\,\mathbf{E}_b\big|\widehat{g}_T(x_0)-(b^j\rho_b)(x_0)\big|^2\Bigg\}\\
&\quad\geq \inf_{\widehat{g}_T}\max\left\{\mathbf{E}_0\left[\psi^{-2}_{\beta_T,L}\left|\widehat{g}_T(x_0)-g_{T,0}(x_0)\right|^2\right],\ \mathbf{E}_1\left[\psi^{-2}_{\beta_*,L}\left|\widehat{g}_T(x_0)-g_{T,1}(x_0)\right|^2\right]\right\}\\
&\quad= \inf_{\widehat{T}_T}\max\left\{\mathbf{E}_0\big|Q_T\widehat{T}_T\big|^2,\ \mathbf{E}_1\big|\widehat{T}_T-\theta_1\big|^2\right\},
\end{aligned}\tag{8.4}$$

where $Q_T := \psi_{\beta_*,L}\psi^{-1}_{\beta_T,L}$, $\widehat{T}_T := \psi^{-1}_{\beta_*,L}\big(\widehat{g}_T(x_0)-g_{T,0}(x_0)\big)$, and

$$\theta_1 := \psi^{-1}_{\beta_*,L}\left(g_{T,1}(x_0)-g_{T,0}(x_0)\right).$$

The proof of the following lemma is completely along the lines of the proof of Theorem 6(i) in Tsybakov [23].

Lemma 4 (Theorem 6(i) in Tsybakov [23]) *Consider Q_T, $\widehat{T}_T$ and θ_1 as introduced above, and assume that $\theta_1\in\mathbb{R}$ satisfies*

$$\big|\theta_1\big|\geq 1-\delta_0. \tag{A1}$$

If $\mathbf{P}_0,\mathbf{P}_1$ are such that $\mathbf{P}_0\ll\mathbf{P}_1$ and, for $\tau>0$ and $\alpha\in(0,1)$ fixed,

$$\mathbf{P}_1\left(\frac{\mathrm{d}\mathbf{P}_0}{\mathrm{d}\mathbf{P}_1}\geq\tau\right)\geq 1-\alpha, \tag{A2}$$

then

$$\inf_{\widehat{T}_T}\max\left\{\mathbf{E}_0\big|Q_T\widehat{T}_T\big|^2,\ \mathbf{E}_1\big|\widehat{T}_T-\theta_1\big|^2\right\}\geq\frac{(1-\alpha)\tau(1-2\delta_0)^2\big(Q_T\delta_0\big)^2}{(1-2\delta_0)^2+\tau\big(Q_T\delta_0\big)^2},$$

where the infimum is taken over all $\widehat{T}_T = \psi^{-1}_{\beta_*,L}\big(\widehat{g}_T(x_0)-g_{T,0}(x_0)\big)$.

We proceed with verifying (A1) and (A2). Note first that

$$\Big|\frac{1}{2}\sum_{k=1}^{d} a_{jk}\partial_k g_{T,\beta_*}(x_0)\Big| = L h_{T,\beta_*}^{\beta_*-d/2}\overline{K}_{\beta_*}(0) \overset{(K3)}{\geq} L h_{T,\beta_*}^{\beta_*-d/2}\,(1-\delta_0/2)\;K^*_{\beta_*}(0)$$
$$= (1-\delta_0/2)\,L\left(\frac{d^2 a_{jj}\rho_{T,0}(x_0)\log T}{\beta_*(2\beta_*-d)L^2T}\right)^{\frac{\beta_*-d/2}{2\beta_*}}\frac{2\beta_*}{d}\,\mathbb{I}_{\beta_*}$$
$$= (1-\delta_0/2)\;C_j(\beta_*,L;\rho_{T,0},\sigma)\,\psi_{T,\beta_*}.$$

Since, for T large enough, $\psi_{\beta_*,L}^{-1}\int_{\mathbb{R}^d} g_{T,\beta_*}(y)\mathrm{d}y \leq \delta_0/2$, this implies

$$\big|\theta_1\big| \geq \psi_{\beta_*,L}^{-1}\left|\frac{1}{2}\sum_{k=1}^{d} a_{jk}\left(\partial_k g_{T,\beta_*}(x_0) - \partial_k\rho_{T,0}(x_0)\int_{\mathbb{R}^d} g_{T,\beta_*}(y)\mathrm{d}y\right)\right|$$
$$\geq 1-\delta_0. \tag{8.5}$$

Denote by Y the solution of the SDE $\mathrm{d}Y_t = b_{T,1}(Y_t)\mathrm{d}t + \sigma\,\mathrm{d}W_t$. In order to verify (A2), note that the specifications on pp. 296–297 in Liptser and Shiryaev [14] imply that the likelihood ratio under $\mathbf{P}_1$ is given by

$$\frac{\mathrm{d}\mathbf{P}_0}{\mathrm{d}\mathbf{P}_1}(Y^T) = \frac{\rho_{T,0}}{\rho_{T,1}}(Y_0)\exp\Bigg(-\frac{1}{2}\int_0^T (b_{T,0}-b_{T,1})^\top(Y_u)a^{-1}(b_{T,0}-b_{T,1})(Y_u)\mathrm{d}u$$
$$+\int_0^T\left(\sigma^{-1}\left(b_{T,0}-b_{T,1}\right)\right)^\top(Y_u)\mathrm{d}W_u\Bigg). \tag{8.6}$$

To proceed, set

$$M_t := \int_0^t\left(\sigma^{-1}\left(b_{T,0}-b_{T,1}\right)\right)^\top(Y_u)\mathrm{d}W_u,\quad t\geq 0,$$

denote $g := (b_{T,0}-b_{T,1})^\top a^{-1}(b_{T,0}-b_{T,1})$, and consider the following stationary sequence of random variables,

$$Z_k := \int_{(k-1)t}^{kt} g(Y_u)\mathrm{d}u,\quad k\geq 1.$$

Since $g\in L^1(\mathbf{P}_1)$, it follows from the ergodic theorem that, for any $t>0$,

$$\frac{1}{n}\sum_{k=1}^{n} Z_k = \frac{1}{n}\int_0^{nt} g(Y_u)\mathrm{d}u = \frac{1}{n}\langle M\rangle_{nt} \xrightarrow{\text{a.s.}}_{n\to\infty} tc,$$

where

$$c := \mathbf{E}_1\left[(b_{T,0}-b_{T,1})^{\top}(Y_0)a^{-1}(b_{T,0}-b_{T,1})(Y_0)\right]$$
$$= \mathbf{E}_1\left\|\sigma^{-1}(b_{T,0}-b_{T,1})(Y_0)\right\|^2. \tag{8.7}$$

In particular, this implies by means of the martingale CLT that, for some standard Brownian motion W,

$$\frac{M_{nt}}{\sqrt{n}} \overset{\mathbf{P}_1}{\Longrightarrow}_{n\to\infty} \sqrt{c}\, W_t. \tag{8.8}$$

Denoting by $[s]$ the integer part of s and considering an arbitrary sequence $\gamma(s) \to_{s\to\infty} 0$, it holds

$$\gamma(s)\int_{[s]}^{s} g(Y_u)\mathrm{d}u \overset{\mathbf{P}_1}{\longrightarrow}_{s\to\infty} 0.$$

Choosing $t \equiv 1$ in (8.8) and passing to the continuous-time case, we obtain

$$\frac{M_T}{\sqrt{T}} \overset{\mathbf{P}_1}{\Longrightarrow}_{T\to\infty} Z \sim \mathcal{N}(0,c).$$

It is verified by straightforward algebra that the definition of the hypotheses $b_{T,0}$ and $b_{T,1}$ entails that

$$2\sigma^{-1}\left(b_{T,0}-b_{T,1}\right) = \sigma^{\top}\nabla\left(\log\frac{\rho_{T,0}}{\rho_{T,1}}\right) = \frac{g_{T,\beta_*}\sigma^{\top}\nabla\left(\log\rho_{T,0}\right)+\sigma^{\top}\nabla g_{T,\beta_*}}{\rho_{T,1}}.$$

The definition of g_{T,β_*} further implies that, using in particular (8.1),

$$\begin{aligned}
\mathbf{E}_1\left[\frac{\left\|\sigma^{\top}\nabla g_{T,\beta_*}(Y_0)\right\|^2}{4\rho_{T,1}^2(Y_0)}\right] &= \int_{\mathbb{R}^d}\frac{\left\|\sigma^{\top}\nabla g_{T,\beta_*}(y)\right\|^2}{4\rho_{T,1}(y)}\mathrm{d}y = a_{jj}\int_{\mathbb{R}^d}\frac{\left(\partial_j g_{T,\beta_*}(y)\right)^2}{4\rho_{T,1}(y)}\mathrm{d}y\\
&= a_{jj}^{-1}L^2\, h_{T,\beta_*}^{2\beta_*-d}\int_{\mathbb{R}^d}\overline{K}_{\beta_*}^2\left(\frac{y-x_0}{h_{T,\beta_*}}\right)\frac{\mathrm{d}y}{\rho_{T,1}(y)}\\
&= a_{jj}^{-1}L^2\, h_{T,\beta_*}^{2\beta_*}\int_{\mathbb{R}^d}\overline{K}_{\beta_*}^2(y)\mathrm{d}y\,\frac{\left(1+o_T(1)\right)}{\rho_{T,1}(x_0)}\\
&\le (1-\delta_0/2)^2\, a_{jj}^{-1}L^2 h_{T,\beta_*}^{2\beta_*}\frac{\left(1+o_T(1)\right)}{\rho_{T,1}(x_0)}.
\end{aligned}$$

The last inequality follows from (K1). Thus, plugging in the definition of h_{T,β_*} [see (8.2)] and using (8.3),

$$\mathbf{E}_1\left[\frac{\left\|\sigma^\top \nabla g_{T,\beta_*}(Y_0)\right\|^2}{4\rho_{T,1}^2(Y_0)}\right] \leq (1-\delta_0/2)^2 L^2 \left(\frac{d\rho_{T,0}(x_0)a_{jj}\log T}{\beta_* L^2 T}\right) \frac{\left(1+o_T(1)\right)}{a_{jj}\rho_{T,1}(x_0)}$$
$$\leq \left(1-\delta_0^2/4\right)^2 \frac{d\log T}{T\beta_*}\left(1+o_T(1)\right).$$

It can be shown by analogous arguments that the terms

$$\mathbf{E}_1\left[\frac{g_{T,\beta_*}^2(Y_0)\|\sigma^\top\nabla(\log\rho_{T,0})(Y_0)\|^2}{\rho_{T,1}^2(Y_0)}\right]$$

and

$$\mathbf{E}_1\left[\frac{g_{T,\beta_*}(Y_0)\left(\nabla(\log\rho_{T,0})(Y_0)\right)^\top a\nabla g_{T,\beta_*}(Y_0)}{\rho_{T,1}^2(Y_0)}\right]$$

are asymptotically negligible. Thus, for c defined in (8.7) and whenever δ_0 is small and T is large enough,

$$c \leq \left(1-\delta_0^2/4\right)^2 \frac{d}{\beta_*}\frac{\log T}{T}\left(1+o_T(1)\right).$$

Consequently, for $\tau := \exp\left(-\frac{(1-\delta_0^2/4)d\log T}{2\beta_*}\right)$, it holds a.s.

$$\frac{\log\tau - \log\rho_{T,0}(Y_0) + \log\rho_{T,1}(Y_0) + \frac{1}{2}\langle M\rangle_T}{\sqrt{Tc}} \leq -\frac{\delta_0^2}{8}\sqrt{\frac{d\log T}{\beta_*}} + o_T(1) \to -\infty.$$

The verification of (A2) is accomplished by means of a tightness argument. Consider some sequence of probability measures $(\mathbf{P}_n)_{n\geq 1}$ on some measurable space, converging weakly to some probability measure $\mathbf{P}$. Tightness of $\mathbf{P}_n$ implies that, for any sequence $\gamma_n \to -\infty$,

$$\lim_{m\to\infty} \max\left\{\mathbf{P}\left((-\infty,\gamma_m)\right),\ \sup_{n\in\mathbb{N}} \mathbf{P}_n\left((-\infty,\gamma_m)\right)\right\} = 0.$$

Thus, $\lim_{m\to\infty}\sup_{n\in\mathbb{N}}\mathbf{P}_n\left((-\infty,\gamma_m)\right) = 0$ and $\lim_{m\to\infty}\inf_{n\in\mathbb{N}}\mathbf{P}_n\left((-\infty,\gamma_m)\right) = 1$. In particular,

$$\lim_{m\to\infty}\mathbf{P}_m\left((\gamma_m,\infty)\right) = 1.$$

In the current framework, this last assertion implies that, for

$$\gamma_T := \frac{\log\tau - \log\rho_{T,0}(Y_0) + \log\rho_{T,1}(Y_0) + \frac{1}{2}\langle M\rangle_T}{\sqrt{Tc}} \to -\infty,$$

one has [plugging in (8.6)]

$$\begin{aligned}\lim_{T\to\infty}\mathbf{P}_1\left(\frac{\mathrm{d}\mathbf{P}_0}{\mathrm{d}\mathbf{P}_1}\geq\tau\right) &= \lim_{T\to\infty}\mathbf{P}_1\left(\frac{\rho_{T,0}}{\rho_{T,1}}(Y_0)\exp\left(M_T-\frac{1}{2}\langle M\rangle_T\right)\geq\tau\right)\\ &= \lim_{T\to\infty}\mathbf{P}_1\left(\frac{M_T}{\sqrt{Tc}}\geq\gamma_T\right)=1.\end{aligned}$$

For large enough T and fixed $\tau > 0$ (where $\delta_0 \in (0, 1)$ can be chosen arbitrarily small), we thus obtain

$$\mathbf{P}_1\left(\frac{\mathrm{d}\mathbf{P}_0}{\mathrm{d}\mathbf{P}_1}\geq\tau\right)\geq 1-\delta_0. \tag{8.9}$$

(III) Completion of the proof. In view of (8.5) and (8.9), Lemma 4 gives

$$\begin{aligned}&\inf_{\widehat{g}_T}\sup_{(\beta,L)\in\mathcal{B}_T}\sup_{b\in\Pi(c_1,c_2,\sigma)}\sup_{\rho_b\in\Sigma_T(\beta,L)}\mathbf{E}_b\left|\widehat{g}_T(x_0)-(b^j\rho_b)(x_0)\right|^2\psi_{\beta,L}^{-2}\\ &\overset{(8.4)}{\geq}\inf_{\widehat{T}_T}\max\left\{\mathbf{E}_0\left|Q_T\widehat{T}_T\right|^2,\ \mathbf{E}_1\left|\widehat{T}_T-\theta_1\right|^2\right\}\\ &\geq\frac{(1-\delta_0)\tau(1-2\delta_0)^2\left(Q_T\delta_0\right)^2}{(1-2\delta_0)^2+\tau\left(Q_T\delta_0\right)^2}.\end{aligned}$$

Since, for $C := C_j(\beta_*, L; \rho_{T,0}, \sigma)/C_j(\beta_T, L; \rho_{T,0}, \sigma)$,

$$Q_T=\frac{\psi_{\beta_*,L}}{\psi_{\beta_T,L}}=C\exp\left(-\frac{d}{4\beta_*\beta_T}(\beta_*-\beta_T)(\log T-\log\log T)\right),$$

we have

$$\tau Q_T^2=C^2\exp\left(\frac{d\left(\delta_0^2\beta_T/4-\beta_*\right)}{2\beta_*\beta_T}\log T\right)\times\exp\left(\frac{d\,(\beta_*-\beta_T)}{2\beta_*\beta_T}\log\log T\right).$$

As $T \to \infty$, $\tau Q_T^2 \to \infty$. Choosing $\delta_0 > 1/A$ for A large enough to ensure $\delta_0 < 1/2$, it holds

$$\frac{(1-\delta_0)\tau(1-2\delta_0)^2\left(Q_T\delta_0\right)^2}{(1-2\delta_0)^2+\tau\left(Q_T\delta_0\right)^2}=\frac{(1-\delta_0)(1-2\delta_0)^2\delta_0^2}{\frac{(1-2\delta_0)^2}{\tau Q_T^2}+\delta_0^2}\to_{T\to\infty}(1-\delta_0)(1-2\delta_0)^2.$$

Taking now $A \to \infty$, the assertion follows. □

B.2: Upper bound

Proof (of Theorem 2) Let $\beta \in [\beta_*, \beta_T]$, $L \in \left[L_*, L^*\right]$, $\beta' \in (d/2, \beta]$, $c_1 \in (0, \infty]$, $c_2 > 0$, σ some nondegenerate $\mathbb{R}^{d\times d}$-matrix, $L' > 0$, and fix $j \in \{1, \ldots, d\}$.

Denote by γ_{Ti}, $i \in \mathbb{N}$, functions of T such that $\lim_{T\to\infty} \gamma_{Ti} = 0$. For $\psi_{T,\beta}$ and $C_j(\beta, L; \rho_b, \sigma)$ introduced in (3.1) and (3.3), respectively, recall that $\psi_{\beta,L} = \psi^j_{\beta,L} = \psi_{T,\beta} C_j(\beta, L; \rho_b, \sigma)$. To enlighten notation, the dependence on the coordinate j again will be mostly suppressed. Denote by $\widetilde{T}(\beta)$ the effective noise level under adaptation, defined as

$$\widetilde{T}(\beta) = \widetilde{T}^j(\beta) := \left(\frac{d\rho_b(x_0) a_{jj} \log T}{\beta T}\right)^{1/2}.$$

Consider the following deterministic counterparts of the bandwidth $\widehat{h}_{T,\beta'}$ and the thresholding sequence $\widehat{\eta}_{T,\beta'}$,

$$h_{T,\beta'} := \left(\frac{d\rho_b(x_0) a_{jj} \log T}{\beta' T}\right)^{1/2\beta'} = \widetilde{T}(\beta')^{1/\beta'} \tag{8.10}$$

and

$$\eta_{T,\beta'} := \left(\frac{d\rho_b(x_0) a_{jj} \log T}{\beta' T}\right)^{\frac{\beta'-d/2}{2\beta'}} \left\|K_{\beta'}\right\|_{L^2(\mathbb{R}^d)} = h_{T,\beta'}^{\beta'-d/2} \left\|K_{\beta'}\right\|_{L^2(\mathbb{R}^d)}.$$

Set

$$\widetilde{\beta} = \widetilde{\beta}(\beta, \beta') := \begin{cases} \beta' + \frac{d}{2}, & \text{if } \frac{d}{2} \le \beta' \le \frac{\beta}{2} + \frac{d}{4}, \\ \beta, & \text{if } \frac{\beta}{2} + \frac{d}{4} < \beta' \le \beta. \end{cases}$$

Define $\overline{\delta}_T := (\log T)^{-1}$, and introduce the random event

$$A_{T,\beta'} := \left\{ \left| \left(\widehat{h}^j_{T,\beta'} / h_{T,\beta'}\right)^{\widetilde{\beta}'-d/2} - 1 \right| \le \overline{\delta}_T \right\}$$

and the associated deterministic set $\mathcal{H}_{T,\beta'} = \mathcal{H}^j_{T,\beta'} := \left\{ h : \left| \left(h/h_{T,\beta'}\right)^{\widetilde{\beta}'-d/2} - 1 \right| \le \overline{\delta}_T \right\}$. Note that there exists a positive constant c_0 such that

$$\mathcal{H}_{T,\beta'} \subset \left\{ h : \left| \left(h/h_{T,\beta'}\right) - 1 \right| \le c_0 \overline{\delta}_T \right\} =: H_{T,\beta'}.$$

Denote the kernel estimator of $(b^j \rho_b)(x_0)$ with deterministic bandwidth $h \in \mathcal{H}_{T,\beta'}$ by

$$g_{T,\beta'}(x_0, h) := \frac{1}{Th^d} \int_0^T K_{\beta'}\left(\frac{X_u - x_0}{h}\right) \mathrm{d}X^j_u, \tag{8.11}$$

and set $g_{T,\beta'}(x_0) := g_{T,\beta'}\big(x_0, h_{T,\beta'}\big)$. Define

$$\mathrm{s}_T(\beta) := h_{T,\beta}^{-d/2}\sqrt{\frac{\rho_b(x_0)a_{jj}}{T}}\;\|K_\beta\|_{L^2(\mathbb{R}^d)},$$

and let $d_T(\beta) := \sqrt{(d\log T)/\beta}$ such that $\mathrm{s}_T(\beta)d_T(\beta) = \eta_{T,\beta}$. For $\beta' \le \beta$, introduce the auxiliary sequence

$$\tau_T(\beta') := \mathrm{s}_T(\beta')\left(\sqrt{d_T^2(\beta') - d_T^2(\beta)} + \left(\frac{\log T}{\beta_T}\right)^{1/4}\right).$$

Lemma 5 (Bound on the bias) *Consider the estimator $g_{T,\beta'}(x_0, h)$ defined in (8.11), and let*

$$\mathrm{b}_{\beta,\beta'} := \left(\frac{2\beta' - d}{d}\right)^{\frac{d/2-\widetilde{\beta}}{2\beta'}}\left((2\pi)^{-d}\int_{\mathbb{R}^d}\frac{\|\lambda\|^{4\beta'-2\widetilde{\beta}}}{\left(1+\|\lambda\|^{2\beta'}\right)^2}\,\mathrm{d}\lambda\right)^{1/2}.$$

For any $\mathcal{H}_{T,\beta'} \ni h > 0$,

$$\sup_{b\in\widetilde{\Pi}(c_1,c_2,\sigma)}\ \sup_{\rho_b\in\Sigma_T(\beta,L)}\left|\mathbf{E}_b g_{T,\beta'}(x_0,h) - (b^j\rho_b)(x_0)\right| \le L h^{\widetilde{\beta}-d/2}\mathrm{b}_{\beta,\beta'}. \quad (8.12)$$

Furthermore,

$$\sup_{d/2<\beta'\le\beta<\infty}\mathrm{b}_{\beta,\beta'} < \infty, \qquad \limsup_{\delta\to 0}\ \sup_{\beta,\beta'\in[\beta_*,\infty):\ |\beta-\beta'|\le\delta}\frac{\mathrm{b}_{\beta,\beta'}}{\mathrm{b}_{\beta,\beta}} \le 1.$$

Proof The bound on the bias in (8.12) is proven by standard arguments and relies in particular on exploiting the scaling properties of the Fourier transform of $K_{\beta',h}$. The remaining assertions are Lemma 1(ii), (iii) in Klemelä and Tsybakov [11]. □

The principal importance of exponential bounds on the stochastic error of estimators considered in the adaptive procedure was already indicated in the introduction. The Bernstein-type deviation inequality (BI) and its implication (BI+), the basic uniform exponential inequality stated in Lemma 2, can be applied to derive more specific bounds on the stochastic error of the estimators $g_{T,\beta}$ defined according to (8.11). The following function classes are defined analogously to Butucea [3],

$$\mathcal{K}_1 := \left\{K_{\beta',h}(\cdot) := h^{-d}K_{\beta'}((\cdot - x_0)/h) \mid h \in H_{T,\beta'}\right\},$$
$$\mathcal{K}_2 := \left\{K_{\beta',h} - K_{\beta',h_{T,\beta'}} \mid h \in H_{T,\beta'}\right\}.$$

For $h \in H_{T,\beta}$, let

$$Z_{T,\beta}(h) := g_{T,\beta}(x_0,h) - \mathbf{E}_b g_{T,\beta}(x_0,h) \\ = \frac{1}{T}\int_0^T K_{\beta,h}(X_u)\mathrm{d}X_u^j - \int_{\mathbb{R}^d} K_{\beta,h}(y)(b^j\rho_b)(y)\mathrm{d}y. \tag{8.13}$$

Lemma 6 *Grant Assumptions* (BI) *and* (**SG+**). *For any $\beta' > d/2$, the stochastic error $Z_{T,\beta'}(\cdot)$ defined according to* (8.13) *has the following properties:*

(a) *For any $u \in \left[\tau_T(\beta'),\ R_1\, \mathrm{s}_T(\beta')\sqrt{\log T}\right]$, $R_1 > 0$ an absolute constant, there exist some sufficiently small $\gamma > 0$, independent of β', and some universal constant $c_1' > 0$ such that*

$$\mathbf{P}_b\left(\sup_{h\in H_{T,\beta'}} \left|Z_{T,\beta'}(h)\right| > u\right) \le c_1' \exp\left(-\frac{1}{2}\left(\frac{u(1-\gamma)}{\mathrm{s}_T(\beta')}\right)^2\right) + o\left(T^{-1}\right). \tag{8.14}$$

(b) *For any $u \in \left[R_1\, \mathrm{s}_T(\beta')\sqrt{\log T},\ \ R_2\right]$, $R_1, R_2 > 0$ absolute constants, it holds, for some absolute constants $c_2', c_3' > 0$,*

$$\mathbf{P}_b\left(\sup_{h\in H_{T,\beta'}} \left|Z_{T,\beta'}(h)\right| > u\right) \le c_2' \exp\left(-c_3'\left(\frac{u}{\mathrm{s}_T(\beta')}\right)^2\right).$$

(c) *Assume that $\beta' < \beta$. Then, uniformly in $\beta \in \mathcal{B}_T$,*

$$\sup_{\substack{\beta'\in\mathcal{B},\\ \beta'<\beta}} m \sup_{b\in\widetilde{\Pi}(c_1,c_2)} \sup_{\rho_b\in\Sigma_T(\beta,L)} \psi_{\beta,L}^{-2} \\ \times \mathbf{E}_b\left[\left(\sup_{h\in H_{T,\beta'}}\left|Z_{T,\beta'}(h)\right|\right)^2 \mathbb{1}\left\{\sup_{h\in H_{T,\beta'}}\left|Z_{T,\beta'}(h)\right| > \tau_T(\beta')\right\}\right] \to 0,$$

$$\sup_{\substack{\beta'\in\mathcal{B},\\ \beta'<\beta}} m \sup_{b\in\widetilde{\Pi}(c_1,c_2)} \sup_{\rho_b\in\Sigma_T(\beta,L)} \psi_{\beta,L}^{-2} \\ \times \mathbf{E}_b\left[\left(\sup_{h\in H_{T,\beta'}}\left|Z_{T,\beta'}(h)\right|\right)^2 \mathbb{1}\left\{\sup_{h\in H_{T,\beta'}}\left|Z_{T,\beta'}(h)\right| > \sqrt{\mathrm{s}_T(\beta')\psi_{\beta,L}}\right\}\right] \to 0.$$

Proof The assertions are analogue to the statements in Lemma 4.3, Lemma 4.5 and Theorem 4.6 in Butucea [3]. For deriving the inequality (8.14) with the specific factor $1/2$ in the exponent in the current diffusion framework, we however have to go into greater detail. Throughout the proof, $D_1, D_2, \ldots$ denote positive constants. For fixed $\delta' \in (0,1)$ and arbitrary $u, T > 0$, write

$$\mathbf{P}_b\left(\left|Z_{T,\beta'}(h_{T,\beta'})\right| > u\right) \le \mathsf{t}_1 + \mathsf{t}_2,$$

for

$$\mathsf{t}_1 := \mathbf{P}_b\left(\left|\frac{1}{T}\int_0^T\left(K_{\beta',h_{T,\beta'}}(X_u)b^j(X_u) - \int_{\mathbb{R}^d} K_{\beta',h_{T,\beta'}}(y)(b^j\rho_b)(y)\mathrm{d}y\right)\mathrm{d}u\right| > \delta' u\right),$$

and, denoting $M_t(K) := \int_0^t K(X_u)\sum_{r=1}^d \sigma_{jr}\mathrm{d}W_u^r$, $t > 0$,

$$\mathsf{t}_2 := \mathbf{P}_b\left(\left|\frac{1}{T}M_T\left(K_{\beta',h_{T,\beta'}}\right)\right| > (1-\delta')u\right).$$

For any $u \le R_1\,\mathrm{s}_T(\beta')\sqrt{\log T}$, we have

$$u\left\|K_{\beta,h_{T,\beta'}}\right\|_\infty \le u h_{T,\beta'}^{-d} K_{\max} \le R_1\,\mathrm{s}_T(\beta') h_{T,\beta'}^{-d}\sqrt{\log T}\; K_{\max} \le D_1\,\mathrm{s}_T^2(\beta') T h_{T,\beta'}^{-d/2+\beta'},$$

such that, since $\beta' > d/2$,

$$\frac{u\|K_{\beta,h}\|_\infty}{\mathrm{s}_T^2(\beta')T} = o_T(1).$$

Furthermore, the enhanced spectral gap assumption (**SG+**) gives

$$\varsigma_b\left(K_{\beta,h_{T,\beta'}}\right) \le D_2 \times \begin{cases} 1, & d = 1,\\ \max\left\{1, (\log(h_{T,\beta'}^{-4}))^2\right\}, & d = 2,\\ h_{T,\beta'}^{2-d}, & d \ge 3. \end{cases}$$

Thus, for T sufficiently large, $C_B\left(\varsigma_b\left(K_{\beta,h_{T,\beta'}}\right) + \delta' u\,\left\|K_{\beta,h_{T,\beta'}}\right\|_\infty\right) \le \mathrm{s}_T^2(\beta')T$. The Bernstein-type deviation inequality (BI) therefore implies that

$$\mathsf{t}_1 \le 2\exp\left(-\frac{\delta'^2u^2}{2\,\mathrm{s}_T^2(\beta')}\right). \tag{8.15}$$

For bounding t_2 from above, we first use Bernstein's inequality for continuous martingales which gives, for any $h > 0$,

$$\mathbf{P}_b\left(\left|M_T\left(K_{\beta',h}\right)\right| > T(1-\delta')u;\ \left\langle M\left(K_{\beta',h}\right)\right\rangle_T \le T^2\,\mathrm{s}_T^2(\beta')\right) \le 2\exp\left(-\frac{(1-\delta')^2u^2}{2\,\mathrm{s}_T^2(\beta')}\right). \tag{8.16}$$

By means of (BI) and using again that $\beta' > d/2$, it can be shown that

$$\begin{aligned}\mathbf{P}_b\left(\left\langle M\left(K_{\beta',h_{T,\beta'}}\right)\right\rangle_T > T^2\,\mathrm{s}_T^2(\beta')\right) &= \mathbf{P}_b\left(\frac{1}{T}\int_0^T K^2_{\beta',h_{T,\beta'}}(X_u)\mathrm{d}u > a_{jj}^{-1}h_{T,\beta'}^{-d}\right)\\ &= o\left(T^{-1}\right).\end{aligned} \tag{8.17}$$

Adding the upper bounds (8.15), (8.16) and (8.17), we obtain, for some small $\gamma > 0$,

$$\mathbf{P}_b\Big(\big|Z_{T,\beta'}(h_{T,\beta'})\big| > u\Big) \leq 2\exp\left(-\frac{u^2(1-\gamma)}{2\,\mathrm{s}_T^2(\beta')}\right) + o\big(T^{-1}\big). \tag{8.18}$$

Consider the sequence $\delta_{T1} := \beta_T\overline{\delta}_T\sqrt{\log T} = (\log\log T)^{\delta}(\log T)^{-1/2} \to 0$, and note that, for any $u > 0$,

$$\mathbf{P}_b\Big(\sup_{h\in H_{T,\beta'}}\big|Z_{T,\beta'}(h)\big| > u\Big) \leq \mathbf{P}_b\Big(\sup_{h\in H_{T,\beta'}}\big|Z_{T,\beta'}(h) - Z_{T,\beta'}(h_{T,\beta'})\big| > u\delta_{T1}\Big) + \mathbf{P}_b\Big(\big|Z_{T,\beta'}(h_{T,\beta'})\big| > u(1-\delta_{T1})\Big).$$

Since $u(1-\delta_{T1}) \leq u \leq R_1\,\mathrm{s}_T(\beta')\sqrt{\log T}$, (8.18) gives an upper bound on the latter summand. For T large enough, it further holds

$$\Big[\tau_T(\beta')\delta_{T1},\ R_1\delta_{T1}\,\mathrm{s}_T(\beta')\sqrt{\log T}\Big] \subset \left[\frac{\beta_T\overline{\delta}_T\sqrt{\log T}}{\sqrt{T}h_{T,\beta'}^{d/2}},\ \beta_T\overline{\delta}_T\right].$$

Taking into account that

$$\sup_{h\in H_{T,\beta'}}\big\|K_{\beta',h} - K_{\beta',h_{T,\beta'}}\big\|_{L^2(\mu_b)} \leq O(1)\sup_{h\in H_{T,\beta'}} h_{T,\beta'}^{-d/2}\left|1-\big(h/h_{T,\beta'}\big)^{2\beta'}\right| \leq O(1)h_{T,\beta'}^{-d/2}\,\beta_T\overline{\delta}_T$$

and since

$$\int_0^1 \max\left\{\sqrt{\log N_{[\,]}(\varepsilon,\mathcal{K}_2,L^2(\mu_b))},1\right\}\mathrm{d}\varepsilon \leq D_4\beta_T\overline{\delta}_T\sqrt{\log T}h_{T,\beta'}^{-d/2},$$

the uniform exponential inequality (BI+) implies that

$$\mathbf{P}_b\Big(\sup_{h\in H_{T,\beta'}}\Big|Z_{T,\beta'}(h) - Z_{T,\beta'}(h_{T,\beta'})\Big| > u\delta_{T1}\Big) \leq C_1\exp\left(-\frac{D_5Th_{T,\beta'}^d(u\delta_{T1})^2}{(\beta_T\overline{\delta}_T)^2}\right). \tag{8.19}$$

Summing the upper bounds due to (8.18) and (8.19), we obtain (8.14).

The inequality stated in (b) follows as an application of (BI+) with $K := h_{T,\beta'}^{-d}K_{\max}$ and

$$M^2 := h_{T,\beta'}^{-d}\,\|K_{\beta'}\|_{L^2(\mathbb{R}^d)}^2\,\rho_b(x_0).$$

Finally, part (c) is proven similarly to Theorem 4.6 in Butucea [3] by noting that there exists some positive constant R such that $\sup_{h \in H_{T,\beta'}} \left|Z_{T,\beta'}(h)\right| \leq R h_{T,\beta'}^{-d}$. A suitable decomposition of

$$\mathbf{E}_b\left[\left(\sup_{h \in H_{T,\beta'}} \left|Z_{T,\beta'}(h)\right|\right)^2 \mathbb{1}\left\{\sup_{h \in H_{T,\beta'}} \left|Z_{T,\beta'}(h)\right| > \tau_T(\beta')\right\}\right]$$

and uniform exponential bounds on the corresponding integrands as they follow from parts (a) and (b) of this lemma then yield the assertions. □

The next lemma contains a decomposition of the normalizing factor $\psi_{\beta,L}$ and some relations which are needed later in the proof. For $h_{T,\beta}$ defined according to (8.10), denote

$$\mathrm{b}_{T,\beta'} := L \mathrm{b}_{\beta,\beta'} h_{T,\beta'}^{\widetilde{\beta}-d/2}. \tag{8.20}$$

Lemma 7 *Let $\beta \in [d/2, \infty)$, $L \in \left[L_*, L^*\right]$, and denote $\nu = (\beta, L)$. It then holds*

$$\psi_\nu = L^{d/(2\beta)} \left(\eta_{T,\beta} + h_{T,\beta}^{\beta-d/2} \mathrm{b}_{\beta,\beta}\right). \tag{8.21}$$

Furthermore, there exist positive constants $D_1, \ldots, D_5$, depending only on β_*, L_*, L^*, d and σ, such that

$$D_1 \leq \psi_\nu / \eta_{T,\beta} \leq D_2, \tag{8.22}$$

and, for $\beta' \in [d/2, \infty)$, $\beta' < \beta$,

$$\frac{D_3}{\beta_T} \left(\frac{\log T}{T}\right)^{\kappa(\beta')-\kappa(\beta)} \leq \frac{\psi_{\beta',L}}{\psi_\nu} \leq D_4 T^{\kappa(\beta)-\kappa(\beta')} \tag{8.23}$$

and

$$\frac{\mathrm{b}_{T,\beta'}^2 + \tau_T^2(\beta')}{\psi_\nu^2} \leq D_5 \log T \; T^{2\kappa(\beta)-2\kappa(\beta')}. \tag{8.24}$$

Proof The proof of the decomposition is comparable to the derivation of relation (68) on p. 461 in Klemelä and Tsybakov [11]; for details, see the proof of Lemma 2.6.6 in Strauch [21]. Assertions (8.22) and (8.23) follow analogously to the proof of the relations (44)–(46) in Lemma 4 in Klemelä and Tsybakov [11] (pp. 453–454). For the proof of (8.24), we refer to the proof of Lemma 3.5 in Butucea [3]. □

Main part of the proof of the upper bound. Define $\beta^- = \beta^-(\beta)$ by

$$\beta^- := \beta - \frac{\beta_T^+}{\log T},$$

where $\beta_T^+ := (\log\log T)^{\delta'}$, for some $\delta' \in (\delta, 1)$. We follow the standard approach and decompose the risk successively. Assume that $b \in \widetilde{\Pi}(c_1, c_2, \sigma)$, let $\nu = (\beta, L)$, and set

$$\mathcal{R}_{T,\nu}^+ = \mathcal{R}_{T,\nu}^+(j) := \sup_{\rho_b \in \Sigma_T(\beta,L)} \psi_\nu^{-2}\, \mathbf{E}_b \left[\left| \widetilde{g}_T^j(x_0) - (b^j \rho_b)(x_0) \right|^2 \mathbb{1}\left\{ \widehat{\beta}_T^j \geq \beta^- \right\} \right],$$
$$\mathcal{R}_{T,\nu}^- = \mathcal{R}_{T,\nu}^-(j) := \sup_{\rho_b \in \Sigma_T(\beta,L)} \psi_\nu^{-2}\, \mathbf{E}_b \left[\left| \widetilde{g}_T^j(x_0) - (b^j \rho_b)(x_0) \right|^2 \mathbb{1}\left\{ \widehat{\beta}_T^j < \beta^- \right\} \right].$$

For ease of notation, we usually suppress the dependence of the risk on the coordinate j.

(I) We first consider the case $\widehat{\beta}_T^j \geq \beta^-$, and we show that

$$\limsup_{T\to\infty} \sup_{\nu \in \mathcal{B}_T} \mathcal{R}_{T,\nu}^+ \leq 1. \tag{8.25}$$

Define $\overline{\beta} = \overline{\beta}(\beta)$ via the equation $\left(\frac{\log T}{L^2 T}\right)^{1/(2\beta)} = \left(\frac{\log T}{T}\right)^{1/(2\overline{\beta})}$. Let $\beta^+ \in \mathcal{G}$ be the largest grid point $\leq \overline{\beta}$, and assume that T is large enough for ensuring $\beta^- < \beta^+$. Denote $\mathcal{G}_1 = \mathcal{G}_1(\beta) := \left\{\beta' \in \mathcal{G} \mid \beta^- \leq \beta' \leq \beta^+\right\}$, $\mathcal{G}_2 = \mathcal{G}_2(\beta) := \left\{\beta' \in \mathcal{G} \mid \beta^+ < \beta' \leq \beta_T\right\}$, and rewrite

$$\mathcal{R}_{T,\nu}^+ = \sup_{\rho_b \in \Sigma_T(\beta,L)} \psi_\nu^{-2}\, \mathbf{E}_b \left[\left| \widetilde{g}_T^j(x_0) - (b^j \rho_b)(x_0) \right|^2 \mathbb{1}\left\{ \widehat{\beta}_T^j \in \mathcal{G}_1 \cup \mathcal{G}_2 \right\} \right].$$

Let $\beta' \in \mathcal{G}_1 = \mathcal{G}_1(\beta)$ and $\rho_b \in \Sigma_T(\beta, L)$, and assume that T is so large that $\widetilde{\beta}(\beta, \beta') = \beta$. Using Lemma 5, the facts that $\mathcal{H}_{T,\beta'} \subset H_{T,\beta'}$, that $\beta' \leq \overline{\beta}$ and the definition of $\overline{\beta}$, it can be shown that

$$\sup_{h \in \mathcal{H}_{T,\beta'}} \left| \mathbf{E}_b g_{T,\beta'}(x_0, h) - (b^j \rho_b)(x_0) \right| \leq \Lambda(\beta, \beta') L^{d/(2\beta)} h_{T,\beta}^{\beta - d/2} \mathrm{b}_{\beta,\beta'} \left(1 + \bar{\delta}_T\right),$$

where

$$\Lambda(\beta, \beta') := \left(d\rho_b(x_0) a_{jj}\right)^{\frac{\beta - d/2}{2\beta'} - \frac{\beta - d/2}{2\beta}} (\beta')^{-\frac{\beta - d/2}{2\beta'}} \beta^{\frac{\beta - d/2}{2\beta}}.$$

The following arguments are along the lines of the proof of the upper bound in Klemelä and Tsybakov [11] (see pp. 461–463). For any $\beta' \in \mathcal{G}_1$, there exists some positive constant C such that

$$\left|\beta - \beta'\right| \leq C \beta_T^+ (\log T)^{-1}. \tag{8.26}$$

Since $\Lambda(\beta, \beta')$ is uniformly continuous in $\beta, \beta' \in [\beta_*, \infty)$, this implies that $\Lambda(\beta, \beta') \leq 1 + \gamma_{T1}$. Furthermore, for any $\beta' \in \mathcal{G}_1$, $\beta \in [\beta_*, \beta_T]$, it holds $\mathrm{b}_{\beta,\beta'} \leq \mathrm{b}_{\beta,\beta}\, (1 + \gamma_{T2})$. Consequently, for any $\beta' \in \mathcal{G}_1$, $\rho_b \in \Sigma_T(\beta, L)$,

$$\sup_{h \in \mathcal{H}_{T,\beta'}} \left| \mathbf{E}_b g_{T,\beta'}(x_0, h) - (b^j \rho_b)(x_0) \right| \leq L^{d/(2\beta)} h_{T,\beta}^{\beta - d/2} \mathrm{b}_{\beta,\beta}\, (1 + \gamma_{T3}).$$

Similar arguments (also see the derivation of line (54) on p. 1591 in Klemelä and Tsybakov [10]) yield

$$\begin{aligned}\eta_{T,\beta^+} &\le \left(\frac{4d\rho_b(x_0)a_{jj}}{\beta^+}\right)^{\frac{\beta^+-d/2}{2\beta^+}} \left(\frac{\log T}{T}\right)^{\frac{\overline{\beta}-d/2}{2\overline{\beta}}} \|K_{\beta^+}\|_{L^2(\mathbb{R}^d)}\,(1+\gamma_{T4})\\ &\le L^{d/(2\beta)}\eta_{T,\beta}\,(1+\gamma_{T5}).\end{aligned} \tag{8.27}$$

Recall the definition of the stochastic error $Z_{T,\beta}(\cdot)$. Whenever $\widehat{\beta}_T^j = \beta' \in \mathcal{G}_1$ and the event $A_{T,\beta'}$ holds, the above arguments imply that

$$\begin{aligned}\left|\widetilde{g}_T^j(x_0) - (b^j\rho_b)(x_0)\right| &= \left|\widehat{g}_{T,\beta'}(x_0) - (b^j\rho_b)(x_0)\right|\\ &\le \sup_{h\in\mathcal{H}_{T,\beta'}} \left|g_{T,\beta'}^j(x_0,h) - (b^j\rho_b)(x_0)\right|\\ &\le \sup_{h\in\mathcal{H}_{T,\beta'}} \left|Z_{T,\beta'}(h)\right| + L^{d/(2\beta)}h_{T,\beta}^{\beta-d/2}\mathrm{b}_{\beta,\beta}\,(1+\gamma_{T3}) \qquad (8.28)\\ &\le \sup_{h\in H_{T,\beta'}} \left|Z_{T,\beta'}(h)\right| + \psi_\nu(1+\gamma_{T3}). \qquad (8.29)\end{aligned}$$

The last line holds true since $\mathcal{H}_{T,\beta'} \subset H_{T,\beta'}$ and in view of the decomposition of the normalizing factor ψ_ν according to (8.21). If $\widehat{\beta}_T^j \ge \beta^+$, the definition of the estimator $\widehat{\beta}_T^j$ according to (4.10) implies that $\left|\widehat{g}^j_{T,\widehat{\beta}_T^j}(x_0) - \widehat{g}^j_{T,\beta^+}(x_0)\right| \le \widehat{\eta}_{T,\beta^+}$.
Therefore, if $\widehat{\beta}_T^j = \beta' \in \mathcal{G}_2$, it holds on A_{T,β^+},

$$\begin{aligned}&\left|\widetilde{g}_T^j(x_0) - (b^j\rho_b)(x_0)\right|\\ &\le \left(\left|\widehat{g}^j_{T,\beta'}(x_0) - \widehat{g}^j_{T,\beta^+}(x_0)\right| + \left|\widehat{g}^j_{T,\beta^+}(x_0) - (b^j\rho_b)(x_0)\right|\right)\\ &\le \widehat{\eta}_{T,\beta^+} + \left|\widehat{g}^j_{T,\beta^+}(x_0) - (b^j\rho_b)(x_0)\right|\\ &\le \sup_{h\in\mathcal{H}_{T,\beta^+}} \Big\{\eta_{T,\beta^+}\left(h/h_{T,\beta}\right)^{\beta-d/2} + \left|g_{T,\beta^+}(x_0,h) - \mathbf{E}_b g_{T,\beta^+}(x_0,h)\right|\\ &\quad + \left|\mathbf{E}_b g_{T,\beta^+}(x_0,h) - (b^j\rho_b)(x_0)\right|\Big\}\\ &\overset{(8.28)}{\le} \eta_{T,\beta^+}\left(1+\overline{\delta}_T\right) + \sup_{h\in H_{T,\beta^+}} \left|Z_{T,\beta^+}(h)\right| + L^{d/(2\beta)}h_{T,\beta}^{\beta-d/2}\mathrm{b}_{\beta,\beta}\,(1+\gamma_{T3})\\ &\overset{(8.27)}{\le} \sup_{h\in H_{T,\beta'}} \left|Z_{T,\beta^+}(h)\right| + \left(L^{d/(2\beta)}\eta_{T,\beta}\,(1+\gamma_{T6}) + L^{d/(2\beta)}h_{T,\beta}^{\beta-d/2}\mathrm{b}_{\beta,\beta}(1+\gamma_{T3})\right).\end{aligned}$$

In view of the decomposition (8.21), this last line implies that

$$\left|\widetilde{g}_T^j(x_0) - (b^j\rho_b)(x_0)\right| \mathbb{1}\{\widehat{\beta}_T^j = \beta' \in \mathcal{G}_2\}\, \mathbb{1}\{A_{T,\beta^+}\} \leq \sup_{h\in H_{T,\beta'}} \left|Z_{T,\beta^+}(h)\right| + \psi_\nu\,(1+\gamma_{T7}). \quad (8.30)$$

Thus, using (8.29) and (8.30),

$$\begin{aligned}
&\psi_\nu^{-2}\,\mathbf{E}_b\left[\left|\widetilde{g}_T^j(x_0) - (b^j\rho_b)(x_0)\right|^2 \mathbb{1}\{\widehat{\beta}_T^j \in \mathcal{G}_1\cup\mathcal{G}_2\}\right]\\
&\leq \sum_{\beta'\in\mathcal{G}_1} \mathbf{E}_b\left[\left(1+\gamma_{T3}+\psi_\nu^{-1}\sup_{h\in H_{T,\beta'}}\left|Z_{T,\beta'}(h)\right|\right)^2 \mathbb{1}\{\widehat{\beta}_T^j=\beta'\}\,\mathbb{1}\{A_{T,\beta'}\}\right]\\
&\quad + \sum_{\beta'\in\mathcal{G}_1} \mathbf{E}_b\left[\psi_\nu^{-2}\left|\widetilde{g}_T^j(x_0) - (b^j\rho_b)(x_0)\right|^2 \mathbb{1}\{\widehat{\beta}_T^j=\beta'\}\,\mathbb{1}\{A^{\mathrm{c}}_{T,\beta'}\}\right]\\
&\quad + \sum_{\beta'\in\mathcal{G}_2} \mathbf{E}_b\left[\left(1+\gamma_{T7}+\psi_\nu^{-1}\sup_{h\in H_{T,\beta^+}}\left|Z_{T,\beta^+}(h)\right|\right)^2 \mathbb{1}\{\widehat{\beta}_T^j=\beta'\}\,\mathbb{1}\{A_{T,\beta^+}\}\right]\\
&\quad + \sum_{\beta'\in\mathcal{G}_2} \mathbf{E}_b\left[\psi_\nu^{-2}\left|\widetilde{g}_T^j(x_0) - (b^j\rho_b)(x_0)\right|^2 \mathbb{1}\{\widehat{\beta}_T^j=\beta'\}\,\mathbb{1}\{A^{\mathrm{c}}_{T,\beta^+}\}\right]\\
&=: \sum_{\beta'\in\mathcal{G}_1}\left(\mathsf{p}_1(\beta')+\mathsf{p}_2(\beta')\right) + \sum_{\beta'\in\mathcal{G}_2}\left(\mathsf{p}_3(\beta')+\mathsf{p}_4(\beta')\right), \text{ say.}
\end{aligned}$$

The terms $\mathsf{p}_1(\cdot),\ldots,\mathsf{p}_4(\cdot)$ are now considered separately. Note first that, for any $\beta'\in\mathcal{G}_1$,

$$\begin{aligned}
\mathsf{p}_1(\beta') \leq{}& \mathbf{E}_b\left[\left(1+\gamma_{T3}+\psi_\nu^{-1}\sup_{h\in H_{T,\beta'}}\left|Z_{T,\beta'}(h)\right|\right)^2 \mathbb{1}\left\{\sup_{h\in H_{T,\beta'}}\left|Z_{T,\beta'}(h)\right| > \sqrt{\mathrm{s}_T(\beta')\psi_\nu}\right\}\right]\\
&+\left(1+\gamma_{T3}+\sqrt{\mathrm{s}_T(\beta')\psi_\nu^{-1}}\right)^2 \mathbf{P}_b\left(\widehat{\beta}_T^j=\beta'\right). \qquad (8.31)
\end{aligned}$$

Since (8.26) holds for any $\beta'\in\mathcal{G}_1$, it can be shown by means of (8.22) and (8.23) that

$$\begin{aligned}
\frac{\mathrm{s}_T(\beta')}{\psi_\nu} &\leq \frac{D_1^{-1}D_4\exp\left(C\beta_T^+\right)\,\mathrm{s}_T(\beta')}{\eta_{T,\beta'}} \leq \frac{D_1^{-1}D_4\exp\left(C\beta_T^+\right)}{d_T(\beta_T)}\\
&\leq D_1^{-1}D_4\exp\left(C\beta_T^+\right)\sqrt{\frac{\beta_T}{\log T}} =: A_T.
\end{aligned}$$

The summand in (8.31) is bounded from above by the sum of the terms

$$2\,(1+\gamma_{T3})^2\ \mathbf{P}_b\left(\sup_{h\in H_{T,\beta'}}\left|Z_{T,\beta'}(h)\right| > \sqrt{\mathrm{s}_T(\beta')\psi_\nu}\right) \stackrel{(8.14)}{\leq} 2c_1'(1+\gamma_{T3})^2\exp\left(-\frac{\psi_\nu(1-\gamma_T)^2}{2\,\mathrm{s}_T(\beta')}\right)$$

and

$$2\psi_\nu^{-2}\mathbf{E}_b\left[\left(\sup_{h\in H_{T,\beta'}}\left|Z_{T,\beta'}(h)\right|\right)^2 \mathbb{1}\left\{\sup_{h\in H_{T,\beta'}}\left|Z_{T,\beta'}(h)\right| > \sqrt{\mathrm{s}_T(\beta')\psi_\nu}\right\}\right].$$

Part (c) of Lemma 6 entails that the latter term tends to zero, uniformly in $\beta' \in \mathcal{G}_1$. Therefore,

$$\mathsf{p}_1(\beta') \leq \left(1+\gamma_{T3}+\sqrt{A_T}\right)^2 \mathbf{P}_b\left(\widehat{\beta}_T^j = \beta'\right) + O(1)\exp\left(-\frac{(1-\gamma_T)^2}{2A_T}\right) + o_T(1).$$

Recall that the cardinality m of the grid $\mathcal{G}$ satisfies

$$m \leq k_1^{-1}\beta_T(\log T)^{\delta_1} = k_1^{-1}(\log T)^{\delta_1}(\log\log T)^{\delta}. \tag{8.32}$$

By construction, $\beta^+ \in \mathcal{G}_1$, such that $\mathsf{p}_3(\beta')$ is upper-bounded analogously. Consequently,

$$\begin{aligned}\sum_{\beta'\in\mathcal{G}_1}\mathsf{p}_1(\beta') + \sum_{\beta'\in\mathcal{G}_2}\mathsf{p}_3(\beta') &\leq \left(1+\max\left\{\gamma_{T3},\gamma_{T7}\right\}+\sqrt{A_T}\right)^2 \mathbf{P}_b\left(\widehat{\beta}_T^j \in \mathcal{G}_1\cup\mathcal{G}_2\right)\\ &\qquad + O(1)m\exp\left(-\frac{(1-\gamma_T)^2}{2A_T}\right) + o_T(1)\\ &\leq 1+o_T(1).\end{aligned}$$

For $\mathsf{p}_2(\cdot)$ and any $\beta' \in \mathcal{G}_1$, there exists some universal constant c_0 such that

$$\mathsf{p}_2(\beta') \leq \psi_\nu^{-2}\,\mathbf{E}_b\left[\left|\widehat{g}_{T,\beta'}^j(x_0) - (b^j\rho_b)(x_0)\right|^2 \mathbb{1}\left\{A_{T,\beta'}^{\mathrm{c}}\right\}\right] \leq c_0\left(\mathbf{P}_b\left(A_{T,\beta'}^{\mathrm{c}}\right)\right)^{1/2}.$$

The regularity conditions on the bandwidth and the kernel used for defining $\widehat{\rho}_T(x_0)$ ensure that, for any $\beta' \in (d/2,\beta]$ and some sufficiently small constant $\alpha > 0$ fixed,

$$\mathbf{P}_b\left(A_{T,\beta'}^{\mathrm{c}}\right) \leq 2\exp\left(-\frac{T\left((1-\alpha)h_T^d\overline{\delta}_T\rho_T^*\right)^2}{2\|Q\|_\infty^2}\right) = o_T(1). \tag{8.33}$$

This implies that $\mathsf{p}_2(\beta')$ is exponentially small, for any $\beta' \in \mathcal{G}_1$, such that $\sum_{\beta'\in\mathcal{G}_1} \mathsf{p}_2(\beta') \to 0$. Analogously, it follows that $\sum_{\beta'\in\mathcal{G}_2} \mathsf{p}_4(\beta') \to 0$, completing finally the verification of (8.25).

(II) It is proven now that

$$\lim_{T\to\infty} \sup_{\nu\in\mathcal{B}_T} \mathcal{R}^-_{T,\nu} = 0. \tag{8.34}$$

Let $\beta' \in \mathcal{G}_T$, and assume that the event $A_{T,\beta'}$ holds. In view of the definition of the stochastic error $Z_{T,\beta'}$ in (8.13) and taking into account Lemma 5, it holds, whenever $\rho_b \in \Sigma_T(\beta, L)$,

$$\begin{aligned}\left|\widehat{g}^j_{T,\beta'}(x_0) - (b^j\rho_b)(x_0)\right| &= \left|\widehat{g}^j_{T,\beta'}\left(x_0, \widehat{h}_{T,\beta'}\right) - (b^j\rho_b)(x_0)\right|\\ &\leq \sup_{h\in\mathcal{H}_{T,\beta'}} \left\{\left|Z_{T,\beta'}(h)\right| + Lh^{\widetilde{\beta}-d/2}\mathrm{b}_{\beta,\beta'}\right\}.\end{aligned}$$

Then, using the definition of $\mathrm{b}_{T,\beta'}$ in (8.20),

$$\begin{aligned}&\sum_{\beta'\in\mathcal{G},\beta'<\beta^-} \sup_{\rho_b\in\Sigma_T(\beta,L)} \psi_\nu^{-2}\,\mathbf{E}_b\left[\left|\widetilde{g}^j_T(x_0) - (b^j\rho_b)(x_0)\right|^2 \mathbb{1}\{\widehat{\beta}^j_T = \beta'\}\,\mathbb{1}\{A_{T,\beta'}\}\right]\\ &\leq \sum_{\beta'\in\mathcal{G},\beta'<\beta^-} \sup_{\rho_b\in\Sigma_T(\beta,L)} \psi_\nu^{-2}\,\mathbf{E}_b\left[\left(\mathrm{b}_{T,\beta'}(1+\overline{\delta}_T) + \sup_{h\in H_{T,\beta'}} \left|Z_{T,\beta'}(h)\right|\right)^2 \mathbb{1}\{\widehat{\beta}^j_T = \beta'\}\right]\\ &\leq 2(1+\overline{\delta}_T)^2 \sum_{\beta'\in\mathcal{G},\beta'<\beta^-} \sup_{\rho_b\in\Sigma_T(\beta,L)} \psi_\nu^{-2}\mathrm{b}^2_{T,\beta'}\,\mathbf{P}_b\left(\widehat{\beta}^j_T = \beta'\right)\\ &\quad +2 \sum_{\beta'\in\mathcal{G},\beta'<\beta^-} \sup_{\rho_b\in\Sigma_T(\beta,L)} \psi_\nu^{-2}\,\mathbf{E}_b\left[\left(\sup_{h\in H_{T,\beta'}} \left|Z_{T,\beta'}(h)\right|\right)^2 \mathbb{1}\{\widehat{\beta}^j_T = \beta'\}\right]\\ &\leq 2 \sum_{\beta'\in\mathcal{G},\beta'<\beta^-} \sup_{\rho_b\in\Sigma_T(\beta,L)} \psi_\nu^{-2}\left(\mathrm{b}^2_{T,\beta'}\left(1+\overline{\delta}_T\right)^2 + \tau^2_T(\beta')\right)\mathbf{P}_b\left(\widehat{\beta}^j_T = \beta'\right)\\ &\quad +2 \sum_{\beta'\in\mathcal{G},\beta'<\beta^-} \sup_{\rho_b\in\Sigma_T(\beta,L)} \psi_\nu^{-2}\,\mathbf{E}_b\left[\left(\sup_{h\in H_{T,\beta'}} \left|Z_{T,\beta'}(h)\right|\right)^2 \mathbb{1}\left\{\sup_{h\in H_{T,\beta'}} \left|Z_{T,\beta'}(h)\right| > \tau_T(\beta')\right\}\right]\\ &=: g_1(\nu) + g_2(\nu).\end{aligned}$$

The term $g_1(\nu)$ is bounded from above by exploiting the fact that the probability to underestimate the value of β by $\widehat{\beta}^j_T$ substantially is small, whenever $\rho_b \in \Sigma_T(\beta, L)$. Recall that m is the cardinality of the grid $\mathcal{G}$.

Lemma 8 (Probability of undershooting) *Let $\beta \in [\beta_*, \infty)$, $\beta' \in \mathcal{G}$, $\beta' < \beta^-$, $L \in [L_*, L^*]$, and $\nu = (\beta, L)$. Then there exists some constant K, depending only on β_*, L_*, L^*, d and σ such that, for any $b \in \widetilde{\Pi}(c_1, c_2, \sigma)$,*

$$\sup_{\rho_b\in\Sigma_T(\beta,L)} \mathbf{P}_b\left(\widehat{\beta}^j_T = \beta'\right) \leq Km\left(T^{-d/(2\beta')} + o(T^{-1})\right). \tag{8.35}$$

Proof The proof substantially relies on applications of Lemma 2. Since the basic arguments are similar to those used in the proof of Lemma 4.8 in Butucea [3] and the proof of Lemma 5 in Klemelä and Tsybakov [11], we do not include the proof but refer to Strauch [21]. □

Let $\beta \in [\beta_*, \beta_T]$, $\beta' \in \mathcal{G}$, $L \in \left[L_*, L^*\right]$, $\nu = (\beta, L)$. By means of Lemma 8 and using relation (8.24) in Lemma 7, we obtain

$$g_1(\nu) = 2 \sum_{\beta' \in \mathcal{G}, \beta' < \beta^-} \sup_{\rho_b \in \Sigma_T(\beta, L)} \psi_\nu^{-2} \left(\mathrm{b}_{T,\beta'}^2 \left(1 + \overline{\delta}_T\right)^2 + \tau_T^2(\beta') \right) \mathbf{P}_b \left(\widehat{\beta}_T^j = \beta' \right)$$
$$\leq 2 D_5 K m \sum_{\beta' \in \mathcal{G}, \beta' < \beta^-} \log T \; T^{-d/(2\beta) + d/(2\beta')} \left(T^{-d/(2\beta')} + o(T^{-1}) \right).$$

In view of the upper bound on the cardinality m of the grid $\mathcal{G} = \mathcal{G}_T$ in (8.32), it follows that $\lim_{T\to\infty} \sup_{\nu \in \mathcal{B}_T} g_1(\nu) = 0$, and the first assertion in Lemma 6(c) immediately gives $\lim_{T\to\infty} \sup_{\nu \in \mathcal{B}_T} g_2(\nu) = 0$. Note finally that, for any $\beta' \in \mathcal{G}$,

$$\sup_{\rho_b \in \Sigma_T(\beta, L)} \psi_\nu^{-2} \mathbf{E}_b \left[\left| \widehat{g}_{T,\beta'}^j(x_0) - (b^j \rho_b)(x_0) \right|^2 \mathbb{1}\left\{ \widehat{\beta}_T^j = \beta' \right\} \mathbb{1}\left\{ A_{T,\beta'}^{\mathrm{c}} \right\} \right]$$
$$\leq \sup_{\rho_b \in \Sigma_T(\beta, L)} \psi_\nu^{-2} \sqrt{ \mathbf{E}_b \left[\left| \widehat{g}_{T,\beta'}^j(x_0) - (b^j \rho_b)(x_0) \right|^4 \right] \mathbf{P}_b \left(A_{T,\beta'}^{\mathrm{c}} \right) } \lesssim \sqrt{ \mathbf{P}_b \left(A_{T,\beta'}^{\mathrm{c}} \right) }.$$

Consequently, taking into account (8.33), it holds, independent both of β and β',

$$\sum_{\beta' \in \mathcal{G}, \beta' < \beta^-} \sup_{\rho_b \in \Sigma_T(\beta, L)} \psi_\nu^{-2} \mathbf{E}_b \left[\left| \widetilde{g}_T^j(x_0) - (b^j \rho_b)(x_0) \right|^2 \mathbb{1}\left\{ \widehat{\beta}_T^j = \beta' \right\} \mathbb{1}\left\{ A_{T,\beta'}^{\mathrm{c}} \right\} \right]$$
$$= o_T(1),$$

thus completing the proof of (8.34). □

References

1. Aït-Sahalia, Y.: Closed-form likelihood expansions for multivariate diffusions. Ann. Stat. **36**(2), 906–937 (2008)
2. Bakry, D., Cattiaux, P., Guillin, A.: Rate of convergence for ergodic continuous Markov processes: Lyapunov versus Poincaré. J. Funct. Anal. **245**(3), 727–759 (2008)
3. Butucea, C.: Exact adaptive pointwise estimation on Sobolev classes of densities. ESAIM: Prob. Stat. **5**, 1–31 (2001)
4. Cattiaux, P., Chafaï, D., Guillin, A.: Central limit theorems for additive functionals of ergodic Markov diffusions processes. ALEA, Lat. Am. J Probab. Math. Stat. **9**(2), 337–382 (2012)
5. Dalalyan, A.S.: Sharp adaptive estimation of the drift function for ergodic diffusions. Ann. Stat. **33**(6), 2507–2528 (2005)
6. Dalalyan, A.S., Kutoyants, Y.A.: Asymptotically efficient trend coefficient estimation for ergodic diffusion. Math. Methods Stat. **11**, 402–427 (2002)
7. Dalalyan, A.S., Reiß, M.: Asymptotic statistical equivalence for ergodic diffusions: the multidimensional case. Probab. Theory Relat. Fields **137**(1), 25–47 (2007)

8. Guillin, A., Léonard, C., Wu, L., Yao, N.: Transportation-information inequalities for Markov processes. Probab. Theory Relat. Fields **144**(3–4), 669–695 (2009)
9. Itō, K., McKean, H.P.: Diffusion Processes and their Sample Paths. Die Grundlehren der mathematischen Wissenschaften in Einzeldarstellungen, vol. 125. Springer, Berlin (1965)
10. Klemelä, J., Tsybakov, A.B.: Sharp adaptive estimation of linear functionals. Ann. Stat. **29**(6), 1567–1600 (2001)
11. Klemelä, J., Tsybakov, A.B.: Exact constants for pointwise adaptive estimation under the Riesz transform. Probab. Theory Relat. Fields **129**, 441–467 (2004)
12. Lepski, O.V.: One problem of adaptive estimation in Gaussian white noise. Theory Probab. Appl. **35**, 459–470 (1990)
13. Lezaud, P.: Chernoff and Berry–Esséen inequalities for Markov processes. ESAIM: Probab. Stat. **5**, 183–201 (2001)
14. Liptser, R.S., Shiryaev, A.N.: Statistics of Random Processes. General Theory of Applications of Mathematics: Stochastic Modelling and Applied Probability, vol. 1, 2nd edn. Springer, Berlin (2001)
15. Metafune, G., Pallara, D., Rhandi, A.: Global properties of invariant measures. J. Funct. Anal. **223**(2), 396–424 (2005)
16. Parzen, E.: On estimation of a probability density function and mode. Ann. Math. Stat. **33**(3), 1065–1076 (1962)
17. Qian, Z., Zheng, W.: A representation formula for transition probability densities of diffusions and applications. Stoch. Process. Appl. **111**(1), 57–76 (2004)
18. Revuz, D., Yor, M.: Continuous Martingales and Brownian Motion. Grundlehren der mathematischen Wissenschaften, vol. 293, 3rd edn. Springer, Berlin (1999)
19. Royer, G.: An initiation to logarithmic Sobolev inequalities. Collection SMF: Cours spécialisés. American Mathematical Society (2007)
20. Spokoiny, V.G.: Adaptive drift estimation for nonparametric diffusion model. Ann. Stat. **28**(3), 815–836 (2000)
21. Strauch, C.: Sharp adaptive drift estimation and Donsker-type theorems for multidimensional ergodic diffusions. Ph.D. thesis, Universität Hamburg (2013)
22. Stroock, D.W., Varadhan, S.R.S.: Multidimensional Diffusion Processes. Grundlehren der mathematischen Wissenschaften, vol. 233. Springer, Berlin, New York (1979)
23. Tsybakov, A.B.: Pointwise and sup-norm sharp adaptive estimation of functions on the Sobolev classes. Ann. Stat. **26**(6), 2420–2469 (1998)
24. van de Geer, S.: Empirical Processes in M-Estimation. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge (2000)
25. van der Vaart, A.W., Wellner, J.W.: Weak Convergence and Empirical Processes. Springer Series in Statistics. Springer, New York (1996)