Gobet, Emmanuel; Hoffmann, Marc; Reiß, Markus

Working Paper

Nonparametric estimation of scalar diffusions based on low frequency data is ill-posed

SFB 373 Discussion Paper, No. 2002,57

Provided in Cooperation with:
Collaborative Research Center 373: Quantification and Simulation of Economic Processes, Humboldt University Berlin

Suggested Citation: Gobet, Emmanuel; Hoffmann, Marc; Reiß, Markus (2002) : Nonparametric estimation of scalar diffusions based on low frequency data is ill-posed, SFB 373 Discussion Paper, No. 2002,57, Humboldt University of Berlin, Interdisciplinary Research Project 373: Quantification and Simulation of Economic Processes, Berlin, https://nbn-resolving.de/urn:nbn:de:kobv:11-10049155

This Version is available at:
https://hdl.handle.net/10419/65321

Nonparametric estimation of scalar diffusions based on low frequency data is ill-posed

Emmanuel Gobet[1], Marc Hoffmann[2] and Markus Reiß[3]

June 25, 2002

[1]*Ecole Polytechnique,* [2]*Laboratoire de Probabilités et Modèles aléatoires - Université Paris 7 and* [3]*Institut für Mathematik, Humboldt-Universität zu Berlin.*

Abstract

We study the problem of estimating the coefficients of a diffusion $(X_t, t \geq 0)$; the estimation is based on discrete data $X_{n\Delta}, n = 0, 1, \ldots, N$. The sampling frequency Δ^{-1} is constant, and asymptotics are taken as the number of observations tends to infinity. We prove that the problem of estimating both the diffusion coefficient – the volatility – and the drift in a nonparametric setting is ill-posed: The minimax rates of convergence for Sobolev constraints and squared-error loss coincide with that of a respectively first and second order linear inverse problem. To ensure ergodicity and limit technical difficulties we restrict ourselves to scalar diffusions living on a compact interval with reflecting boundary conditions.

An important consequence of this result is that we can characterize quantitatively the difference between the estimation of a diffusion in the low frequency sampling case and inference problems in other related frameworks: nonparametric estimation of a diffusion based on continuous or high frequency data, but also parametric estimation for fixed Δ, in which case $\sqrt{N}$-consistent estimators usually exist. Our approach is based on the spectral analysis of the associated Markov semigroup. A rate-optimal estimation of the coefficients is obtained via the nonparametric estimation of an eigenvalue-eigenfunction pair of the transition operator of the discrete time Markov chain $(X_{n\Delta}, n = 0, 1, \ldots, N)$ in a suitable Sobolev norm, together with an estimation of its invariant density.

2000 Mathematics Subject Classification: 62G99, 62M05, 62M15
Key words and phrases: Diffusion processes, Nonparametric estimation, Discrete sampling, Low frequency data, Spectral approximation, Ill-posed problems.
Acknowledgments. *This work was partially supported by the European network Dynstoch and the DFG-Sonderforschungsbereich 373.*

1 Introduction and main results

1.1 Overview

Since Feller's celebrated classification, stationary scalar diffusions have served as a representative model for homogeneous Markov processes in continuous time. Historically, diffusion processes were probably first seen as approximation models for discrete Markov chains, up to an appropriate rescaling in time and space. More recently, the development of financial mathematics has argued in favour of genuinely continuous time models, with simple dynamics governed by a local mean (drift) $b(\bullet)$ and local variance (diffusion coefficient, or volatility) $\sigma(\bullet)$ on the state space $S = \mathbb{R}$ or $S \subset \mathbb{R}$ with appropriate boundary condition. The dynamics are usually described by an Itô-type stochastic differential equation in the interior of S, which in the time-homogeneous case reads like

$$dX_t = b(X_t)\,dt + \sigma(X_t)\,dW_t, \quad t \geq 0,$$

where the driving process $(W_t,\, t \geq 0)$ is standard Brownian motion. The growing importance of diffusion models progressively raised among the community of statisticians a vast research program, from both quantitative and theoretical angles. We outline the main achievements of this program in Section 1.2 and give some bibliographical references.

In the late 1970 s a statistician was able to characterize qualitatively the properties of a parametric ergodic diffusion model based on the continuous observation of a sample path

$$X^T := (X_t, 0 \leq t \leq T)$$

of the trajectory, as $T \to \infty$, i.e. as the time length of the experiment grows to infinity, a necessary assumption to assure the growing of information thanks to the recurrence of the sample path. The 1980 s explored various discretization schemes of the continuous time model: the data X^T could progressively be replaced by the more realistic observation

$$X^{N,\Delta_N} := (X_{n\Delta_N}, n = 0, 1, \ldots, N),$$

with asymptotics taken as $N \to \infty$. The discretization techniques used at that time required the high frequency sampling assumption $\Delta_N \to 0$ whereas $N\Delta_N \to \infty$ in order to guarantee the closeness of X^{N,Δ_N} and X^T, with $T = N\Delta_N$. Not too long after, a similar nonparametric program was achieved both for continuous time and high frequency data.

By the early up to mid 1990 s, the frontier remained the "fixed Δ case", i.e. the case of low frequency data. This is the topic of the present paper. First, one must understand the importance and flexibility gained by being able to relax the assumption that the sampling time Δ between two data points is "small": Indeed, one can hardly deny that in practice, it may well happen that sampling

with arbitrarily small Δ is simply not feasible. Put differently, the asymptotic statistical theory is a mathematical construct to assess the quality of an estimator based on discrete observations and it must be decided which asymptotics are adequate for the data at hand. Second, the statistical nature of the problem drastically changes when passing from high to low frequency sampling: the approximation properties of the sample path $X^{N\Delta_N}$ by X^{N,Δ_N} are not valid anymore; the observation $(X_0, X_\Delta, \dots, X_{N\Delta})$ becomes a genuine Markov chain, and inference about the underlying coefficients of the diffusion process must be sought via the identification of the law of the observation X^{N,Δ_N}. In the time-homogeneous case the mathematical properties of the random vector X^{N,Δ_N} are embodied in the transition operator

$$P_\Delta f(x) := \mathbb{E}[f(X_\Delta) \,|\, X_0 = x],$$

defined on appropriate test functions f (The symbol $\mathbb{E}$ denotes expecation w.r.t. the probability space on which is properly defined the diffusion X, see below.) Under suitable assumptions, the operator P_Δ is associated with a Feller semigroup $(P_t, t \geq 0)$ with a densely defined infinitesimal generator L on the space of continuous functions given by

$$Lf(x) = L_{\sigma,b} f(x) := \frac{\sigma^2(x)}{2} f''(x) + b(x) f'(x).$$

The second order term $\sigma(\bullet)$ is the diffusion coefficient, and the first-order term $b(\bullet)$ the drift coefficient. Postulating the existence of an invariant density $\mu(\bullet) = \mu_{\sigma,b}(\bullet)$, the operator L is unbounded, but self-adjoint negative on $L^2(\mu) := \{f \mid \int |f|^2 \, d\mu < \infty\}$ and the functional calculus gives the correspondence

$$P_\Delta = \exp(\Delta L) \tag{1.1}$$

in the operator sense. Therefore, a consistent statistical program can be resumed to start from the observed Markov chain $X^{N,\Delta}$, estimate its transition operator P_Δ, and infer about the pair $\big(b(\bullet), \sigma(\bullet)\big)$, via the correspondence (1.1), or, in other words via the spectral properties of the operator P_Δ. Expressed as a diagram, we obtain the following line

$$\texttt{data} = X^{N,\Delta} \xrightarrow{(E)} P_\Delta \xrightarrow{(I)} L \longleftrightarrow \big(b(\bullet), \sigma(\bullet)\big) = \texttt{parameter}. \tag{1.2}$$

The efficiency of a given statistical estimation procedure will be measured by the proficiency in combining the (E) – or estimation – part and the (I) – or identification – part of the model.

The works of Hansen and Scheinkman (1995) and later Hansen, Scheinkman and Touzi (1998) and Chen (1998) paved the way: They formulated a precise and thorough program, proposing and discussing several methods for identifying scalar diffusions through their spectral properties. In particular, their paper of 1998 and the references therein provide a deep understanding of the problem at hand. Simultaneously and independently, the Danish school, impulsed

by the works of Kessler and Sørensen (1999), systematically studied the parametric efficiency of spectral methods in the fixed Δ setting described above. By constructing estimating functions based on eigenfunctions of the operator L, they could construct $\sqrt{N}$-consistent estimators and obtained precise asymptotic properties.

However, a quantitative study of nonparametric estimation in the fixed Δ context remained out of reach for some time, both for technical and conceptual reasons. The purpose of the present paper is to fill in this gap, by trying to understand and explain why the nonparametric case significantly differs from its parametric analogue, as well as from the high frequency data framework in nonparametrics. We are going to establish minimax rates of convergence (in short MRC, see also Section 1.3.1) over various smoothness classes, characterizing upper and lower bounds for estimating $b(\bullet)$ and $\sigma(\bullet)$ based on the obervation of $X_0, X_\Delta, \ldots, X_{N\Delta}$, with asymptotics taken as $N \to \infty$. The MRC is an index of both accuracy of estimation and complexity of the model. We will show that in the nonparametric case, the complexity of the problem of estimating $b(\bullet)$ and $\sigma(\bullet)$ is related to a certain ill-posed inverse problem. Although we mainly focus on the theoretical aspects of the statistical model, the estimators we propose are based on feasible nonparametric smoothing methods: They can be implemented in practice, allowing for adaptivity and finite sample optimisation.

1.2 Statistical estimation for diffusions: an outlook

In this section, we give a brief and selective summary of the evolution of the area over the last two decades. The nonparametric identification of diffusion processes from continuous data was probably first addressed in the reference paper of Banon (1978). More precise estimation results can be listed as follows:

1.2.1 Continuous or high frequency data: the parametric case

Estimation of a finite dimensional parameter θ from $X^T = (X_t, 0 \le t \le T)$ with asymptotics as $T \to \infty$ when X is a diffusion of the form

$$dX_t = b_\theta(X_t)dt + \sigma(X_t)dW_t \tag{1.3}$$

is classical (Brown and Hewitt (1975)). Here $(W_t, t \ge 0)$ is a standard Wiener process. The diffusion coefficient is perfectly identified from the data by means of the quadratic variation of X. By assuming the process X to be ergodic (positively recurrent), a sufficiently regular parametrisation $\theta \mapsto b_\theta(\bullet)$ implies the LAN property (Local Asymptotic Normality, see e.g. Ibragimov and Khas'minskii (1981)) for the underlying statistical model, therefore ensuring the $\sqrt{T}$-consistency and efficiency of the ML-estimator (see Liptser and Shiryaev (2001)).

In the case of discrete data $X_{n\Delta_N}, n = 0, 1, \ldots, N$ with high frequency sampling $\Delta_N^{-1} \to \infty$, but long range observation $N\Delta_N \to \infty$ as $N \to \infty$, various discretization schemes and estimating procedures have been proposed (Dacunha-Castelle and Florens (1986), Yoshida (1992), Kessler (1997)) until

Gobet (2002) eventually proved the LAN-property for ergodic diffusions of the form

$$dX_t = b_{\theta_1}(X_t)dt + \sigma_{\theta_2}(X_t)dW_t \tag{1.4}$$

in a general setting, by means of the Malliavin calculus: Under suitable regularity conditions, the finite-dimensional parameter θ_1 in the drift term can be estimated with optimal rate $\sqrt{N\Delta_N}$, whereas the finite-dimensional parameter θ_2 in the diffusion coefficient is estimated with the optimal rate $\sqrt{N}$.

1.2.2 Continuous or high frequency data: the nonparametric case

A similar program was progressively obtained in nonparametrics: If the drift function $b(\bullet)$ is globally unknown in the model given by (1.3), but belongs to a Sobolev ball $S(s, L)$ (of smoothness order $s > 0$ and radius L) over a given compact interval $\mathcal{I}$, see Appendix 5.2 for a precise definition, a certain kernel estimator $\hat{b}_T(\bullet)$, achieves the following upper bound in $L^2(\mathcal{I})$ and in a root-mean-squared sense:

$$\sup_{b \in S(s,L)} \mathbb{E}\left[\|\hat{b}_T - b\|^2_{L^2(\mathcal{I})}\right]^{1/2} \lesssim T^{-s/(2s+1)}.$$

This already indicates a formal analogy with the model of nonparametric regression or "signal + white noise" where the same rate holds. (Here and in the sequel, the symbol $\lesssim$ means "up to constants", possibly depending on the parameters of the problem, but that are continuous in their arguments.) See Kutoyants (1997), and more recently Dalalyan (2001) for precise results on nonparametric estimation from continuous data.

Similar extensions to the discrete case with high frequency data sampling for the model driven by (1.4) were given in Hoffmann (1999) where the rates $(N\Delta_N)^{-s/(2s+1)}$ for the drift function $b(\bullet)$ and $N^{-s/(2s+1)}$ for the diffusion coefficient $\sigma(\bullet)$ have been obtained and proved to be optimal. See also the pioneering paper of Pham (1981).

1.2.3 Spectral methods for parametric estimation

As already mentioned above, as soon as the sampling frequency $\Delta_N^{-1} = \Delta^{-1}$ is not large anymore, the problem of estimating a parameter in the drift or diffusion coefficient becomes significantly more difficult: the trajectory properties that can be recovered from the data when Δ_N is small are lost; in particular, there is no evident approximating scheme (like, for instance, contrast estimators) that can efficiently compute or mimic the continuous maximum likelihood estimator, which therefore becomes untractable.

Kessler and Sørensen (1999) suggested using eigenvalues λ_θ and eigenvectors $\varphi_\theta(\bullet)$ of the parametrised infinitesimal generator

$$L_\theta f(x) = \frac{\sigma^2_\theta(x)}{2} f''(x) + b_\theta(x) f'(x),$$

i.e., such that $L_\theta \varphi_\theta(x) = \lambda_\theta \varphi_\theta(x)$. Indeed, since the pair $(\lambda_\theta, \varphi_\theta)$ also satisfies

$$P_\Delta \varphi_\theta(X_{n\Delta}) = \mathbb{E}[\varphi_\theta(X_{(n+1)\Delta}) \mid X_{n\Delta}] = \exp(\lambda_\theta \Delta)\varphi_\theta(X_{n\Delta}),$$

whenever easy to compute, the knowledge of a pair $(\lambda_\theta, \varphi_\theta)$ can be translated into a set of conditional moment conditions to be used in estimating functions. With their method, Kessler and Sørensen can construct $\sqrt{N}$-consistent estimators that are nearly efficient. See also the paper of Hansen, Scheinkman and Touzi (1998) that we already mentioned.

In a sense, in this idea lies the essence of our method. However, the strategy of Kessler and Sørensen is not easily extendable to nonparametrics: there is no straightforward way to pass from a finite-dimensional parametrization of the generator L_θ with explicit eigenpairs $(\lambda_\theta, \varphi_\theta)$ to a full nonparametric space with satisfactory approximation properties, like Sobolev balls for instance. Besides, there would be no evident route to treat the variance of such speculative nonparametric estimators either, because the behaviour of the parametric Fisher information matrix for a growing number of parameters is too complex to be easily controlled. We will see in Section 2.2 below how to overpass these objections by estimating directly an eigenpair nonparametrically.

1.2.4 Prospectives

A quick summary yields the following table for optimal rates of convergence:

	Parametric		Nonparametric	
	b	σ	b	σ
Continuous	$T^{-\frac{1}{2}}$	known	$T^{-s/(2s+1)}$	known
HF data	$(N\Delta_N)^{-1/2}$	$N^{-1/2}$	$(N\Delta_N)^{-s/(2s+1)}$	$N^{-s/(2s+1)}$
LF data	$N^{-1/2}$	$N^{-1/2}$	u_N	v_N

This table can be interpreted as follows: the difficulty of the estimation problem is increasing from top to bottom and from left to right. A double horizontal line separates the continuous/high frequency (HF) data domain from the low frequency (LF) data domain. The breach for LF data opened by Kessler and Sørensen as well as Hansen, Scheinkman and Touzi shows that $\sqrt{N}$-consistent estimators usually exist in the parametric case. The remaining case are the rates of convergence for LF data in the nonparametric case u_N for the drift $b(\bullet)$ and v_N for the diffusion coefficient $\sigma(\bullet)$, which we are aiming for.

Another interesting point is that asymptotic estimation of the drift coefficient is never easier than estimation of the diffusion coefficient, a property that will also be kept in the nonparametric LF regime. This is closely connected with the fact that for the convergence of transition densities of diffusions the conditions imposed on the convergence of the drift coefficients are weaker than those for the diffusion coefficients, compare with the monograph by Stroock and Varadhan (1979).

1.3 Nonparametric estimation based on LF data

1.3.1 Rates of convergence in the minimax theory

By passing from the parametric to the nonparametric setting in the table above, we observe a remarkable correspondence between the exponent $1/2$ and the exponent $s/(2s+1)$. Indeed, this phenomenon is well understood in regular Gaussian experiments and can be described as follows. Let us consider the classical "signal + white noise" model: we aim at recovering an unknown signal $f(\bullet)$ from data $\dot{Y}$ obeying

$$\dot{Y} = f + \frac{1}{\sqrt{N}}\dot{W}, \tag{1.5}$$

where $\dot{W}$ is white noise and $N^{-1/2}$ is a – small – noise level[1]. If $f(\bullet) = f_\theta(\bullet)$ is known up to a finite dimensional parameter θ and if the parametrisation $\theta \mapsto f_\theta(\bullet)$ is regular, then the model (1.5) based on the observation of $\dot{Y}$ is regular as well – it satisfies the LAN property – and optimal $\sqrt{N}$-consistent estimators exist. If we relax the parametrization by θ and allow f to lie in a Sobolev ball $S(s, L)$ of smoothness $s > 0$ and radius $L > 0$, then, one can show (e.g. Ibragimov and Khas'minskii (1981)) that certain kernel or projection estimators $\hat{f}(\bullet) = \hat{f}(\bullet, \dot{Y})$ satisfy

$$\sup_{f \in S(s,L)} \mathbb{E}\left[\|\hat{f} - f\|_{L^2([0,1])}\right]^{1/2} \lesssim N^{-s/(2s+1)}, \tag{1.6}$$

The inflation of the risk bound shows a dimension effect governed by the smoothness of the parameter $f(\bullet)$, best understood by looking at projection methods: the variance-bias decomposition reads

$$\hat{f} - f = (\hat{f} - f_m) + (f_m - f),$$

where f_m is the best approximation of f in an m-dimensional approximation space W_m, having accuracy $\|f_m - f\|_{L^2} \lesssim m^{-s}$, while the variance term satisfies

$$\mathbb{E}[\|\hat{f} - f_m\|_{L^2}^2]^{1/2} \lesssim N^{-1/2} \dim(W_m) = N^{-1/2} m. \tag{1.7}$$

The balance between the two terms yields $m \sim N^{1/(2s+1)}$ hence (1.6) (here and in the sequel, the notation $c_1 \sim c_2$ means $c_1 \lesssim c_2$ and $c_2 \lesssim c_1$.) In particular, we see that the rate $\sqrt{N}$ is achievable whenever $f(\bullet)$ is finite-dimensional, i.e. m is kept constant.

[1] In probabilitic terms, we observe

$$Y(dx) = f(x)dx + \frac{1}{\sqrt{N}} W(dx),\ x \in [0,1]$$

where $W(dx)$ is a centred Gaussian measure of intensity the Lebesgue measure dx on the unit square $[0,1]$

Moreover, the upper bound $N^{-s/(2s+1)}$ is optimal:

$$\liminf_{N\to\infty} \inf_{F} \sup_{f\in S(s,L)} N^{s/(2s+1)}\, \mathbb{E}_{b,\sigma} \left[\|F-f\|_{L^2([0,1])}\right]^{1/2} > 0, \tag{1.8}$$

where the infimum in F is taken over all estimators $F = F(\bullet, \dot{Y})$. Putting (1.6) and (1.8) together, we say that the *minimax rate of convergence* (MRC) over Sobolev bodies $S(s,L)$ is (of order) $N^{-s/(2s+1)}$. As already mentioned, it is noteworthy that $N^{-s/(2s+1)}$ is both an index of accuracy and complexity: two different statistical models sharing the same MRC do have common geometric features, see the review article of Nussbaum (1999).

Taking advantage of the parametric results of Kessler and Sørensen, we might be tempted at first sight to conjecture the corresponding results in the nonparametric case based on LF data. But this is not the case.

1.4 Main result

The symmetry we observe between the parametric and nonparametric case is broken for nonparametric estimation based on LF data. Combining Theorem 2.5 and 3.1 below, we see that if $b(\bullet)$ and $\sigma(\bullet)$ are of Sobolev smoothness order s_1 and s_2, under some further conditions the MRC are

$$u_N = N^{-s_1/(2s_1+5)} \text{ for } b(\bullet) \text{ and } v_N = N^{-s_2/(2s_2+3)} \text{ for } \sigma(\bullet).$$

Although different from the classical case, these risk bounds do have an analogue in the Gaussian white noise setting if we move a little beyond the simple white noise heuristics. Instead of (1.5) we consider the statistical model generated by data $\dot{Y}$, with

$$\dot{Y} = Kf + \frac{1}{\sqrt{N}}\dot{W}, \tag{1.9}$$

and where now $K : L^2([0,1]) \to L^2([0,1])$ is a compact linear operator, which has a regularization property of order $\alpha > 0$ in the following sense: The operator K maps L^2 into the Sobolev space H^α continuously. For instance, α-fold integration has regularizing order α. Assuming K^{-1} to exist, we can formally apply this operator and obtain

$$K^{-1}\dot{Y} = f + \frac{1}{\sqrt{N}}K^{-1}\dot{W}, \tag{1.10}$$

which shows that the bias term is the same as for (1.5), but the stochastic variance term increases due to the bad condition of K^{-1} on finite-dimensional spaces. The regularization property leads to the growth $N^{-1/2}m^{\alpha+1/2}$ of this stochastic error in terms of the dimensionality m of an approximation space, which should be compared with the $m^{1/2}$-increase in (1.7).

Classical nonparametric theory (see e.g. Korostelev and Tsybakov (1993)) shows that the minimax rate of convergence over Sobolev balls of smoothness

$s > 0$ is then of order $N^{-s/(2s+2\alpha+1)}$. The inflation in the denominator by 2α is the unavoidable payment for the ill-posedness of the problem. This sheds light on our present problem: estimating $\sigma(\bullet)$ and $b(\bullet)$ appears now as an ill-posed inverse problem of order $\alpha = 1$ and $\alpha = 2$, respectively. As will become apparent in the sequel, this can be understood by the ill-posedness involved in the identification step (I) explained in diagram (1.2). The fact that the estimation rates between $b(\bullet)$ and $\sigma(\bullet)$ differ will become more transparent in Section 2.2 (see also the remarks following Theorem 2.5).

1.5 Organisation of the paper

In Section 2, we give a precise mathematical and statistical framework for estimating the diffusion coefficient $\sigma(\bullet)$ and the drift $b(\bullet)$ in a scalar diffusion model with boundary conditions. Section 2.2 gives the precise construction of nonparametric estimators for $b(\bullet)$, $\sigma(\bullet)$ based on low frequency discrete data by means of spectral methods. Theorem 2.5 in Section 2.3 provides an upper bound in L^2-loss, uniformly over Sobolev bodies for our procedure. A subsequent discussion reviews the advantages and limitations of our assumptions and methodology. Possible extensions to related aspects of nonparametric estimation, like adaptation to unknown smoothness, are discussed. A detailed proof of the upper bound is given in Section 2.4.

The upper bound of Theorem 2.5 is proved to be optimal in Section 3. The investigation of lower bounds for Markov chains has some history, see Birgé (1983). Although the route we follow is classical, the technique we employ is new, and uses genuine Hilbert space method in order to handle an appropriate statistical distance in our model, induced by the operator P_Δ.

Complementary remarks and comments are given in Section 4. An appendix – Section 5 – contains important auxiliary technical results about the spectral properties of scalar diffusions, together with the essential ingredients of approximation theory (e.g., multiresolution analysis, inverse and direct estimates) that we use throughout the paper.

2 Estimation of drift and diffusion coefficient

2.1 A diffusion model with boundary reflections

We shall restrict ourselves to reflecting diffusions on a one-dimensional interval, because their theory is highly developed so that e.g. no further technical ergodicity conditions have to be postulated. Choosing for convenience the interval $[0,1]$, we suppose

2.1. Assumption. *The function $b : [0,1] \to \mathbb{R}$ is measurable and bounded, the function $\sigma : [0,1] \to (0,\infty)$ is continuous and positive and the function $\nu : [0,1] \to \mathbb{R}$ satisfies $\nu(0) = 1$, $\nu(1) = -1$.*

We consider the stochastic differential equation

$$\begin{cases} dX_t = b(X_t)\,dt + \sigma(X_t)\,dW_t + \nu(X_t)\,dL_t(X), \quad X_0 = x_0 \in [0,1] \\ X_t \in [0,1] \quad \forall t \geq 0. \end{cases} \tag{2.1}$$

The process $(W_t,\, t \geq 0)$ is a standard Brownian motion and $(L_t(X),\, t \geq 0)$ is a nonanticipative continuous nondecreasing process that increases only when $X_t \in \{0,1\}$. By Lions and Sznitman (1984), there exists a weak solution of this stochastic differential equation. The boundedness of $b(\bullet)$ and the ellipticity of $\sigma(\bullet)$ even ensure that this solution is strong and unique (e.g. Cepa (1995)). Note that the process $L(X)$ is part of the solution and is given by a difference of local times of X at the boundary points of $[0,1]$.

Due to the compactness of $[0,1]$ and the reflecting boundary conditions, the Markov process X has a so-called spectral gap, which implies a geometrical decay of the covariances $\mathrm{Cov}(X_m, X_{m+n}) \to 0$ for $m \in \mathbb{N}$, $n \to \infty$, see also Lemma 5.1 and Lemma 5.2 in the appendix. In particular, a unique invariant measure μ exists and the one-dimensional distributions of X_n converge exponentially fast to μ as $n \to \infty$ so that the assumption of stationarity can be made without loss of generality for asymptotic results.

We denote by $\mathbb{P}_{\sigma,b}$ the law of the associated stationary diffusion on the canonical space $\Omega = \mathcal{C}(\mathbb{R}_+, [0,1])$ of continuous functions over the positive axis with values in $[0,1]$, equipped with the topology of uniform convergence and endowed with its Borel σ-field $\mathcal{F}$. We denote by $\mathbb{E}_{\sigma,b}$ the corresponding expectation operator. Given $N \geq 1$ and $\Delta > 0$, we observe the canonical process $(X_t, t \geq 0)$ at equidistant times $n\Delta$ for $n = 0, 1, \ldots, N$. Let $\mathcal{F}_N$ denotes the σ-field generated by $X_{n\Delta}, n = 0, \ldots, N$.

2.2. Definition. *An estimator of the pair $\big(\sigma(\bullet), b(\bullet)\big)$ is an $\mathcal{F}_N$-measurable function with values in $L^2([0,1]) \times L^2([0,1])$.*

In order to assess the L^2-risk in a minimax framework, we introduce the nonparametric set Θ_s, which consists of pairs of functions of regularity s and $s-1$, respectively.

2.3. Definition. *For $s > 1$ and given constants $C \geq c > 0$, we consider the class $\Theta(s, C, c)$ defined by*

$$\Big\{ (\sigma, b) \in H^s([0,1]) \times H^{s-1}([0,1]) \;\Big|\; \|\sigma\|_{H^s} \leq C,\, \|b\|_{H^{s-1}} \leq C,\, \inf_x \sigma(x) \geq c \Big\},$$

where H^s denotes the Sobolev space of order s (see Appendix 5.2).

We shall be inexact and write Θ_s for $\Theta(s, C, c)$. Note that all $\big(\sigma(\bullet), b(\bullet)\big) \in \Theta_s$ satisfy Assumption 2.1.

2.2 Construction of the estimators

We shall construct our estimator of the diffusion coefficient $\sigma(\bullet)$ and of the drift coefficient $b(\bullet)$ by applying spectral methods, thereby passing from the transition

operator P_Δ, which is approximately known to us, to the infinitesimal generator L, which more explicitly encodes the functions $\sigma(\bullet)$ and $b(\bullet)$. In the sequel, we shall often be using the specific form of the invariant density

$$\mu(x) = 2C_0\sigma^{-2}(x)\exp\Big(\int_0^x 2b(y)\,\sigma^{-2}(y)\,dy\Big) \tag{2.2}$$

and the function $S(\bullet) = 1/s'(\bullet)$, derived from the scale function $s(\bullet)$,

$$S(x) = \tfrac{1}{2}\sigma^2(x)\mu^{-1}(x) = C_0\exp\Big(\int_0^x 2b(y)\sigma^{-2}(y)\,dy\Big) \tag{2.3}$$

with the normalizing constant $C_0 > 0$, depending on $\sigma(\bullet)$ and $b(\bullet)$, cf. Section VII.3 in Revuz, Yor (1999). The action of the generator in divergence form is given by

$$Lf(x) = L_{\sigma,b}f(x) = \tfrac{1}{2}\sigma^2(x)f''(x) + b(x)f'(x) = \tfrac{1}{\mu(x)}\big(S(x)f'(x)\big)', \tag{2.4}$$

where the domain of this unbounded operator on $L^2(\mu)$ is given by the subspace of the Sobolev space H^2 with Neumann boundary conditions

$$\mathrm{dom}(L) = \{f \in H^2([0,1]) \mid f'(0) = f'(1) = 0\}.$$

The generator L is a selfadjoint elliptic operator on $L^2(\mu)$ with compact resolvent so that it has nonpositive point spectrum only. If ν_1 denotes the largest negative eigenvalue of L with eigenfunction u_1, then due to the reflecting boundary of $[0,1]$ Neumann boundary conditions $u_1'(0) = u_1'(1) = 0$ hold and we find

$$Lu_1 = \mu^{-1}(Su_1')' = \nu_1 u_1 \Longrightarrow S(x)u_1'(x) = \nu_1\int_0^x u_1(y)\mu(y)\,dy. \tag{2.5}$$

From (2.3) we can thus derive an explicit expression for the diffusion coefficient:

$$\sigma^2(x) = \frac{2\nu_1\int_0^x u_1(y)\mu(y)\,dy}{u_1'(x)\mu(x)}. \tag{2.6}$$

The corresponding expression for the drift coefficient is

$$b(x) = \nu_1\frac{u_1(x)u_1'(x)\mu(x) - u_1''(x)\int_0^x u_1(y)\mu(y)\,dy}{u_1'(x)^2\mu(x)}. \tag{2.7}$$

Hence, if we knew the invariant measure μ, the eigenvalue ν_1 and the eigenfunction u_1 (including its first two derivatives), we could exactly determine the drift and diffusion coefficient. Of course, these identities are valid for any eigenfunction u_k with eigenvalue ν_k, but for the sake of concreteness and better numerical stability we shall only use the largest non-degenerate eigenvalue ν_1. Moreover, it is known that only the eigenfunction u_1 does not have a vanishing derivative in the interior of the interval (cf. Proposition 5.5) so that by this choice indeterminacy at interior points is avoided. The main point is now that from

discrete time observations $(X_{n\Delta})$ we can estimate the invariant measure μ and the transition operator P_Δ. Using semigroup theory (Theorem IV.3.7 in Engel and Nagel (2000)), we know that u_1 is also an eigenfunction of P_Δ with eigenvalue $\lambda_1 = e^{\Delta\nu_1}$. Thus, our procedure is to determine estimates $\hat{\mu}$ of μ and $\hat{P}_\Delta$ of P_Δ, to calculate the corresponding eigenpair $(\hat{\lambda}_1, \hat{u}_1)$ and to use (2.6) and (2.7) to build a plug-in estimator of $\sigma(\bullet)$ and $b(\bullet)$.

We are now describing the estimation procedure for the diffusion coefficient $\sigma(\bullet)$ in detail, while the construction for $b(\bullet)$ follows the same lines. As we shall later need to calculate spectral information we shall be using projection methods for nonparametric smoothing, which have the advantage of approximating the abstract operators by finite-dimensional matrix estimators. More specifically, we make use of wavelets on the interval $[0,1]$. In order to be self-contained, we recall the essential ingredients we need in Appendix 5.2 (see also the monograph of Cohen (2000)).

2.4. Definition. *Let (ψ_λ) with multindices $\lambda = (j,k)$ be a compactly supported orthonormal wavelet basis of $L^2([0,1])$. Each $f \in L^2([0,1])$ has the unique expansion*

$$f = \sum_\lambda \langle f, \psi_\lambda \rangle \psi_\lambda.$$

The corresponding approximation spaces (V_J) are defined as L^2-closed linear span of the wavelets up to level $J \geq -1$

$$V_J := \overline{\text{span}}\{\psi_\lambda \,|\, |\lambda| \leq J\}, \qquad \textit{where } |(j,k)| := j.$$

The L^2-orthogonal projection onto V_J is called π_J, the $L^2(\mu)$-orthogonal projection onto V_J is called π_J^μ.

The canonical projection estimate of μ based on $(X_{n\Delta})_{0\leq n\leq N}$ is given by

$$\hat{\mu} := \sum_{|\lambda|\leq J} \hat{\mu}_\lambda \psi_\lambda \tag{2.8}$$

for some $J \in \mathbb{N}$ and with the empirical wavelet coefficients

$$\hat{\mu}_\lambda := \frac{1}{N+1} \sum_{n=0}^{N} \psi_\lambda(X_{n\Delta}). \tag{2.9}$$

By the ergodicity of X (see Section 2.1) it follows that $\hat{\mu}_\lambda$ is a consistent estimate of $\langle \mu, \psi_\lambda \rangle$ for $N \to \infty$. In order to estimate the action of the transition operator on the wavelet basis $(\mathbf{P}_\Delta^J)_{\lambda,\lambda'} := \langle P_\Delta \psi_\lambda, \psi_{\lambda'} \rangle_\mu$, we introduce the symmetrized matrix estimator $\hat{\mathbf{P}}_\Delta$:

$$(\hat{\mathbf{P}}_\Delta)_{\lambda,\lambda'} := \frac{1}{2N} \sum_{n=1}^{N} \Big(\psi_\lambda(X_{(n-1)\Delta}) \psi_{\lambda'}(X_{n\Delta}) + \psi_{\lambda'}(X_{(n-1)\Delta}) \psi_\lambda(X_{n\Delta}) \Big). \tag{2.10}$$

Note that in this case, $\mathbb{P}_{\sigma,b}$-almost surely and in $L^2(\mathbb{P}_{\sigma,b})$:

$$(\hat{\mathbf{P}}_\Delta)_{\lambda,\lambda'} \to \mathbb{E}_{\sigma,b}[\psi_\lambda(X_0)\psi_{\lambda'}(X_\Delta)] = \langle \psi_\lambda, P_\Delta \psi_{\lambda'} \rangle_\mu \text{ for } N \to \infty.$$

We thus merely obtain an approximation of the transition operator with respect to the unknown scalar product $\langle \bullet, \bullet \rangle_\mu$ in $L^2(\mu)$. We therefore introduce a third statistic, which approximates the $|V_J| \times |V_J|$-dimensional Gram matrix $\mathbf{G}$ with entries $\mathbf{G}_{\lambda,\lambda'} = \langle \psi_\lambda, \psi_{\lambda'} \rangle_\mu$, given by:

$$\hat{\mathbf{G}}_{\lambda,\lambda'} := \frac{1}{N}\Big(\tfrac{1}{2}\psi_\lambda(X_0)\psi_{\lambda'}(X_0) + \tfrac{1}{2}\psi_\lambda(X_{N\Delta})\psi_{\lambda'}(X_{N\Delta}) + \\ + \sum_{n=1}^{N-1} \psi_\lambda(X_{n\Delta})\psi_{\lambda'}(X_{n\Delta}) \Big). \tag{2.11}$$

The particular treatment of the boundary terms will be explained later. Note that the matrix $\hat{\mathbf{G}}$ is indeed a consistent estimator of $\mathbf{G}$. If we put $\Sigma = (\kappa_n \psi_\lambda(X_n))_{|\lambda| \le J,\, n \le N}$ with $\kappa_0 = \kappa_N = \frac{1}{2}$ and $\kappa_n = 1$ otherwise, we have $\hat{\mathbf{G}} = N^{-1}\Sigma\Sigma^T$, where Σ^T denotes the transpose of Σ. Our construction can thus be regarded as a least-squares type estimator, like in a usual regression setting, see the argument developed in (2.14) below.

We combine the last two estimators in order to determine estimates for the eigenvalue λ_1 and the eigenfunction u_1 of P_Δ. First, as will be made precise in Proposition 2.12, the operators P_Δ and $\pi_J^\mu P_\Delta$ are close for large values of J. Note that all eigenvectors of $\pi_J^\mu P_\Delta$ lie in V_J, the range of $\pi_J^\mu P_\Delta$. Since the constant functions are eigenfunctions of both, P_Δ and $\pi_J^\mu P_\Delta$, the second largest eigenvalues λ_1 and λ_1^J, respectively, satisfy the variational characterisation (see Section 4.5 in Davies (1996)):

$$\lambda_1 = \sup_{\substack{\|f\|_\mu = 1 \\ \langle f, \mathbf{1} \rangle_\mu = 0}} \langle P_\Delta f, f \rangle_\mu, \qquad \lambda_1^J = \sup_{\substack{f \in V_J,\, \|f\|_\mu = 1 \\ \langle f, \mathbf{1} \rangle_\mu = 0}} \langle P_\Delta f, f \rangle_\mu.$$

This shows immediately $\lambda_1^J \uparrow \lambda_1$, i.e. the true eigenvalue is approximated by λ_1^J, but always underestimated. The eigenvalue λ_1^J and its eigenfunction u_1^J are characterized by

$$\langle P_\Delta u_1^J, \psi_\lambda \rangle_\mu = \lambda_1^J \langle u_1^J, \psi_\lambda \rangle_\mu \qquad \forall\, |\lambda| \le J. \tag{2.12}$$

We pass to vector-matrix notation and use from now on bold letters to define for a function $v \in V_J$ the corresponding coefficient column vector $\mathbf{v} = (\langle v, \psi_\lambda \rangle)_{|\lambda| \le J}$. Observe carefully the different scalar products in $L^2([0,1])$ used; here it is with respect to the Lebesgue measure. Thus, we can rewrite (2.12) as

$$\mathbf{P}_\Delta^J \mathbf{u}_1^J = \lambda_1^J \mathbf{G} \mathbf{u}_1^J. \tag{2.13}$$

As $\mathbf{v}^T \mathbf{G} \mathbf{v} = \langle v, v \rangle_\mu > 0$ holds for $v \in V_J \setminus \{0\}$, the matrix $\mathbf{G}$ is invertible and $(\lambda_1^J, \mathbf{u}_1^J)$ is an eigenpair of the matrix $\mathbf{G}^{-1}\mathbf{P}_\Delta^J$. Observe that this matrix is selfadjoint with respect to the scalar product induced by $\mathbf{G}$:

$$\langle \mathbf{G}^{-1}\mathbf{P}_\Delta^J \mathbf{v}, \mathbf{w} \rangle_{\mathbf{G}} := (\mathbf{G}^{-1}\mathbf{P}_\Delta^J \mathbf{v})^T \mathbf{G} \mathbf{w} = \mathbf{v}^T \mathbf{P}_\Delta^J \mathbf{w} = \langle \mathbf{v}, \mathbf{G}^{-1}\mathbf{P}_\Delta^J \mathbf{w} \rangle_{\mathbf{G}}.$$

Similarly, $\mathbf{v}^T\hat{\mathbf{G}}\mathbf{v} = N^{-1}(\Sigma^T\mathbf{v})^T\Sigma^T\mathbf{v} \geq 0$ holds and the matrix $\hat{\mathbf{G}}$ can be shown to be even strictly positive definite with high probability (see Lemma 2.19). In this case, we obtain the corresponding symmetry property

$$\langle \hat{\mathbf{G}}^{-1}\hat{\mathbf{P}}_\Delta \mathbf{v}, \mathbf{w}\rangle_{\hat{\mathbf{G}}} = (\hat{\mathbf{P}}_\Delta \mathbf{v})^T\mathbf{w} = \langle \mathbf{v}, \hat{\mathbf{G}}^{-1}\hat{\mathbf{P}}_\Delta \mathbf{w}\rangle_{\hat{\mathbf{G}}}.$$

The Cauchy-Schwarz-inequality and the inequality between geometric and arithmetic mean yield the estimate

$$\begin{aligned}
\langle \hat{\mathbf{G}}^{-1}\hat{\mathbf{P}}_\Delta \mathbf{v}, \mathbf{v}\rangle_{\hat{\mathbf{G}}} &= \frac{1}{N}\sum_{n=1}^{N} v(X_{(n-1)\Delta})v(X_{n\Delta}) \\
&\leq \frac{1}{N}\left(\sum_{n=0}^{N-1} v(X_{n\Delta})^2\right)^{1/2}\left(\sum_{n=1}^{N} v(X_{n\Delta})^2\right)^{1/2} \\
&\leq \frac{1}{N}\left(\tfrac{1}{2}v(X_0)^2 + \tfrac{1}{2}v(X_{N\Delta})^2 + \sum_{n=1}^{N-1} v(X_{n\Delta})^2\right) \\
&= \langle \mathbf{v}, \mathbf{v}\rangle_{\hat{\mathbf{G}}}.
\end{aligned}$$

We infer that all eigenvalues of $\hat{\mathbf{G}}^{-1}\hat{\mathbf{P}}_\Delta$ are real and not larger than one. The eigenvalue one is even attained by the eigenvector corresponding to the constant function $\mathbf{1}$. Hence, the second largest eigenvalue $\hat{\lambda}_1$ of $\hat{\mathbf{G}}^{-1}\hat{\mathbf{P}}_\Delta$ is well defined and not larger than one. For this purpose we had chosen to downweight the boundary terms of $\hat{\mathbf{G}}$. Then the eigenvalue $\hat{\lambda}_1$ of $\hat{\mathbf{G}}^{-1}\hat{\mathbf{P}}_\Delta$ and its corresponding eigenvector $\hat{\mathbf{u}}_1$ yield reasonable estimates of λ_1^J and u_1^J.

The estimator matrix $\hat{\mathbf{G}}^{-1}\hat{\mathbf{P}}_\Delta$ is built as in the least squares approach for projection methods in classical regression. In order to estimate $P_\Delta\psi_{\lambda_0}(x) = \mathbb{E}_{\sigma,b}[\psi_{\lambda_0}(X_\Delta)\,|\,X_0 = x]$, the least squares method consists of minimizing

$$\sum_{n=1}^{N}\Big|\psi_{\lambda_0}(X_n) - \sum_{|\lambda|\leq J}\alpha_\lambda^0\psi_\lambda(X_{n-1})\Big|^2 \longrightarrow \min! \tag{2.14}$$

over all real coefficients (α_λ^0), leading to the normal equations

$$\sum_{n=1}^{N}\Big(\sum_{|\lambda|\leq J}\alpha_\lambda^0\psi_\lambda(X_{n-1})\Big)\psi_{\lambda'}(X_{n-1}) = \sum_{n=1}^{N}\psi_{\lambda'}(X_{n-1})\psi_{\lambda_0}(X_n) \qquad \forall\,|\lambda'| \leq J.$$

Up to the special treatment of the boundary terms, we thus obtain the vector (α_λ^0) as the column with index λ_0 in $\hat{\mathbf{G}}^{-1}\hat{\mathbf{P}}_\Delta$. Thus, our method estimates the coefficients for all basis functions ψ_{λ_0}, $|\lambda_0| \leq J$, simultaneously. See also the model selection approach in nonlinear regression (Baraud (2000) and Baraud, Comte and Viennet (2001)) which is natural in this context, and the final remarks in Section 4.

Plugging the estimator $\hat{\mu}$ and the just defined estimators $\hat{\lambda}_1$ and $\hat{u}_1$ into (2.6) and (2.7), we obtain our estimators of $\sigma^2(\bullet)$ and $b(\bullet)$.

Finally, our estimator of $\big(\sigma(\bullet), b(\bullet)\big)$ is defined by:

$$\widehat{\sigma^2}(x) := \frac{2\Delta^{-1}\log(\hat{\lambda}_1)\int_0^x \hat{u}_1(y)\hat{\mu}(y)\,dy}{\hat{u}_1'(x)\hat{\mu}(x)}, \tag{2.15}$$

$$\hat{b}(x) := \Delta^{-1}\log(\hat{\lambda}_1)\frac{\hat{u}_1(x)\hat{u}_1'(x)\hat{\mu}(x) - \hat{u}_1''(x)\int_0^x \hat{u}_1(y)\hat{\mu}(y)\,dy}{\hat{u}_1'(x)^2\hat{\mu}(x)}. \tag{2.16}$$

To avoid indeterminacy, the denominator of the estimators is forced to remain above a certain minimal level, which depends on the subinterval $[a,b] \subset [0,1]$ forwhich the loss function is taken, see Theorem 2.5 and (2.21).

2.3 Asymptotic upper bound

We can now state the main theorem, giving an asymptotic upper bound for the risk of our estimators as $N \to \infty$ uniformly over the class Θ_s introduced in Definition 2.3. In order to avoid problems due to the vanishing denominator in (2.15), (2.16) at the boundary, we restrict to a proper subinterval of $[0,1]$.

2.5. Theorem. *For all $s > 1$, $C \geq c > 0$ and $0 < a < b < 1$ we have*

$$\limsup_{N\to\infty} N^{s/(2s+3)} \sup_{(\sigma,b)\in\Theta_s} \mathbb{E}_{\sigma,b}\Big[\|\widehat{\sigma^2} - \sigma^2\|^2_{L^2([a,b])}\Big]^{1/2} < \infty$$

and

$$\limsup_{N\to\infty} N^{(s-1)/(2s+3)} \sup_{(\sigma,b)\in\Theta_s} \mathbb{E}_{\sigma,b}\Big[\|\hat{b} - b\|^2_{L^2([a,b])}\Big]^{1/2} < \infty.$$

2.6. Remark. *Comparing with Theorem 3.1 below, upper and lower bounds agree, therefore our estimators are rate-optimal in a minimax sense.*

2.7. Remark. *If we set $s_1 = s$ and $s_2 = s_1 - 1$, we obtain the rates announced in Section 1.4, namely $N^{-s_1/(2s_1+3)}$ for $\sigma(\bullet)$ and $N^{-s_2/(2_s+5)}$ for $b(\bullet)$. More precisely, one can make the link with the diagram and with ill-posed problems by saying that estimation of $\mu(\bullet)$ is well-posed ($N^{-s/(2s+1)}$), but for $S(\bullet)$ we need an estimate of the derivative $u_1'(\bullet)$ yielding a degree one of ill-posedness ($N^{-s/(2s+3)}$). Observe that the regularity conditions $\sigma \in H^s$ and $b \in H^{s-1}$ are translated into $\mu \in H^s$, $S \in H^s$. The transformation of (μ, S) to $\sigma^2(\bullet) = 2S(\bullet)/\mu(\bullet)$ is stable (L^2-continuous for $S(\bullet) \geq s_0 > 0$), whereas in $b(\bullet) = S'(\bullet)/\mu(\bullet)$ another ill-posed operation (differentiation) occurs with degree 1. A brief stepwise explanation reads as follows:*

Step 1: The natural parametrisation (μ, P_Δ) is well-posed (for P_Δ in the strong operator convergence sense). Step 2: The calculation of the spectral pair (λ_1, u_1) is well-posed. Step 3: The differentiation of u_1 that determines S has an ill-posedness of degree 1. Step 4: The calculation of σ^2 from (μ, S) is well-posed. Step 5: The calculation of b from (μ, S) has an ill-posedness of degree 1.

2.4 Proof of the upper bound

2.4.1 Convergence of $\hat{\mu}$

In order to be self-contained, we first recall the proof for the classical risk bound in estimating the invariant measure:

2.8. Proposition. *With the choice $2^J \sim N^{1/(2s+1)}$ the following uniform risk estimate holds for $\hat{\mu}$ based on N observations:*

$$\sup_{(\sigma,b)\in\Theta_s} \mathbb{E}_{\sigma,b}[\|\hat{\mu}-\mu\|_{L^2}^2]^{1/2} \lesssim N^{-s/(2s+1)}.$$

Proof. The explicit formula (2.2) for μ shows that $\|\mu\|_{H^s}$ is uniformly bounded over Θ_s. This implies that the bias term satisfies

$$\|\mu-\pi_J\mu\|_{L^2} \lesssim 2^{-Js}\|\mu\|_{H^s} \sim N^{-s/(2s+1)},$$

uniformly over Θ_s. Since $\hat{\mu}_\lambda$ is an unbiased estimator of $\langle\mu,\psi_\lambda\rangle$, we can apply the variance estimates of Lemma 5.2 to obtain

$$\mathbb{E}_{\sigma,b}[\|\hat{\mu}-\pi_J\mu\|_{L^2}^2] = \sum_{|\lambda|\le J} \mathrm{Var}_{\sigma,b}[\hat{\mu}_\lambda] \lesssim 2^J N^{-1},$$

which – in combination with the uniformity of the constants involved – gives the announced upper bound. □

2.4.2 Spectral approximation

We shall rely on the spectral approximation results given in the book of Chatelin (1983), compare also Kato (1995). Since for $\hat{\sigma}(\bullet)$ we have to estimate not only the eigenvalue u_1, but also its derivative u_1', we will be working in the L^2-Sobolev space H^1. The general idea is that the error in the eigenvalue and in the eigenfunction can be controlled by the error of the operator on the eigenspace, once the overall error measured in the operator norm is small. Denoting by $R(T,z) = (T - z\,\mathrm{Id})^{-1}$ the resolvent map of the operator T, by $\sigma(T)$ its spectrum and by $B(x,r)$ the closed ball of radius r around x, the precise statement is given in the following proposition and its corollary:

2.9. Proposition. *Suppose a linear operator T on a Hilbert space has a simple eigenvalue λ such that $\sigma(T)\cap B(\lambda,\rho) = \{\lambda\}$ holds for some $\rho > 0$. Let T_ε be a second linear operator with $\|T_\varepsilon - T\| < \frac{R}{2}$, where $R := (\sup_{z\in B(\lambda,\rho)}\|R(T,z)\|)^{-1}$. Then the operator T_ε has a simple eigenvalue λ_ε in $B(\lambda,\rho)$ and there are eigenvectors u and u_ε with $Tu = \lambda u$, $T_\varepsilon u_\varepsilon = \lambda_\varepsilon u_\varepsilon$ satisfying*

$$\|u_\varepsilon - u\| \le \sqrt{8}R^{-1}\|(T_\varepsilon - T)u\|. \tag{2.17}$$

Proof. We use the resolvent identity and the Cauchy integral representation of the spectral projection P_ε on the eigenspace of T_ε contained in $B(\lambda,\rho)$, see

Lemma 6.4 in Chatelin (1983). By the usual Neumann series argument we find formally for an eigenvector u corresponding to λ

$$\begin{aligned}\|u - P_\varepsilon u\| &= \frac{1}{2\pi}\left\| \oint_{B(\lambda,\rho)} \frac{R(T_\varepsilon, z)}{\lambda - z}\, dz\, (T_\varepsilon - T)u \right\| \\ &\le \frac{1}{2\pi} 2\pi\rho \sup_{z\in B(\lambda,\rho)} \|R(T_\varepsilon, z)\|\rho^{-1}\|(T_\varepsilon - T)u\| \\ &\le \sup_{z\in B(\lambda,\rho)} \frac{\|R(T,z)\|}{1 - \|R(T,z)\|\|T_\varepsilon - T\|}\|(T_\varepsilon - T)u\| \\ &= (R - \|T_\varepsilon - T\|)^{-1}\|(T_\varepsilon - T)u\|.\end{aligned}$$

Hence, for $\|T_\varepsilon - T\| < \frac{R}{2}$ this calculation is a posteriori justified and simplifies further:

$$\|u - P_\varepsilon u\| < 2R^{-1}\|(T_\varepsilon - T)u\|.$$

Applying $\|(T_\varepsilon - T)u\| < \frac{R}{2}\|u\|$ once again, we see that the projection P_ε cannot be zero. Consequently there must be a part of the spectrum of T_ε in $B(\lambda, \rho)$. By the argument in the proof of Theorem 5.22 in Chatelin (1983) this part consists of a simple eigenvalue λ_ε.

It remains to find eigenvectors that are close, too. Observe that for arbitrary Hilbert space elements g, h with $\|g\| = \|h\| = 1$ and $\langle g, h\rangle \ge 0$

$$\|g - h\|^2 = 2 - 2\langle g, h\rangle \le 2(1 + \langle g, h\rangle)(1 - \langle g, h\rangle) = 2\|g - \langle g, h\rangle h\|^2$$

holds, i.e. the distance of the vectors can be estimated by the distance between one vector and its orthogonal projection on the other vector. We substitute for g and h the normalized eigenvectors u and u_ε with $\langle u, u_\varepsilon\rangle \ge 0$, note that oblique projections only enlarge the right hand side and thus infer (2.17). □

2.10. Corollary. *Under the conditions of Proposition 2.9 the eigenvalues λ and λ_ε are also close; more precisely, there is a constant $C = C(R, \|T\|)$ such that*

$$|\lambda_\varepsilon - \lambda| \le C\|(T_\varepsilon - T)u\|.$$

Proof. The inverse triangle inequality yields

$$\begin{aligned}|\lambda_\varepsilon - \lambda| &= |\|T_\varepsilon u_\varepsilon\| - \|Tu\|| \\ &\le \|T_\varepsilon(u_\varepsilon - u) + (T_\varepsilon - T)u\| \\ &\le (\|T\| + \|T_\varepsilon - T\|)\|u_\varepsilon - u\| + \|(T_\varepsilon - T)u\| \\ &\le \big((\|T\| + \tfrac{R}{2})\sqrt{8}R^{-1} + 1\big)\|(T_\varepsilon - T)u\|,\end{aligned}$$

where the last line follows from Proposition 2.9. □

2.4.3 Bias estimates

In a first estimation step, we bound the deterministic error due to the finite dimensional projection $\pi_J^\mu P_\Delta$ of P_Δ. We start with a lemma stating that π_J^μ and π_J have similar approximation properties.

2.11. Lemma. *Let $m : [0,1] \to [m_0, m_1]$ be a measurable function with $m_1 \geq m_0 > 0$. Denote by π_J^m the $L^2(m)$-orthogonal projection onto the multiresolution space V_J. Then there is a constant $C = C(m_0, m_1)$ such that*

$$\|(\mathrm{Id} - \pi_J^m) f\|_{H^1} \leq C \|(\mathrm{Id} - \pi_J) f\|_{H^1} \qquad \forall\, f \in H^1([0,1]).$$

Proof. The norm equivalence $m_0 \|g\|_{L^2} \leq \|g\|_m \leq m_1 \|g\|_{L^2}$ implies

$$\|\pi_J^m\|_{L^2 \to L^2} \leq m_1 m_0^{-1} \|\pi_J^m\|_{L^2(m) \to L^2(m)} = m_1 m_0^{-1}.$$

On the other hand, the Bernstein inequality in V_J and the Jackson inequality for $\mathrm{Id} - \pi_J$ in H^1 and L^2 (see Appendix 5.2) yield for $f \in H^1$

$$\begin{aligned}
\|(\mathrm{Id} - \pi_J^\mu) f\|_{H^1} &= \|(\mathrm{Id} - \pi_J^\mu)(\mathrm{Id} - \pi_J) f\|_{H^1} \\
&\leq \|(\mathrm{Id} - \pi_J) f\|_{H^1} + \|\pi_J^\mu (\mathrm{Id} - \pi_J) f\|_{H^1} \\
&\lesssim \|(\mathrm{Id} - \pi_J) f\|_{H^1} + 2^J \|\pi_J^\mu (\mathrm{Id} - \pi_J) f\|_{L^2} \\
&\leq \|(\mathrm{Id} - \pi_J) f\|_{H^1} + \|\pi_J^\mu\|_{L^2 \to L^2} \|(\mathrm{Id} - \pi_J) f\|_{H^1},
\end{aligned}$$

where the constants only depend on the multiresolution analysis. □

By the preceding lemma and some results on the transition densities, we obtain the first precise approximation result.

2.12. Proposition. *We have uniformly over Θ_s*

$$\|\pi_J^\mu P_\Delta - P_\Delta\|_{H^1 \to H^1} \lesssim 2^{-Js}.$$

Proof. We denote by p_Δ the transition density. Being the kernel of the operator P_Δ, it satisfies

$$P_\Delta f(x) := \int_{[0,1]} f(y) p_\Delta(x,y) dy$$

for any bounded Borel function f. Since $p_\Delta \in H^{s+1,s}([0,1] \times [0,1])$ holds with uniform norm bound on Θ_s by Lemma 5.7, it follows that $P_\Delta : H^1 \to H^{s+1}$ is continuous, also with a uniform norm bound over Θ_s. Thus, by Lemma 2.11 we find

$$\|(P_\Delta - \pi_J^\mu P_\Delta) f\|_{H^1} \lesssim \|\mathrm{Id} - \pi_J\|_{H^{s+1} \to H^1} \|f\|_{H^1}.$$

The Jackson inequality in H^1 (see Appendix 5.2) yields the asserted result. □

2.13. Corollary. *Let λ_1^J be the largest eigenvalue smaller than one of π_J^μ with eigenfunction u_1^J. Then uniformly over Θ_s the following estimate holds:*

$$|\lambda_1^J - \lambda_1| + \|u_1^J - u_1\|_{H^1} \lesssim 2^{-Js}.$$

Proof. We are going to apply Proposition 2.9 on the space H^1 and its Corollary 2.10. In view of the preceding Proposition 2.12, it remains to establish the existence of uniformly strictly positive values for ρ and R over Θ_s. The uniform separation of λ_1 from the rest of the spectrum is the content of Proposition 5.5 in Appendix 5.1.

For the choice of ρ in Proposition 5.5 we search a uniform bound R. If we regard P_Δ on $L^2(\mu)$, then P_Δ is selfadjoint and satisfies $\|R(P_\Delta, z)\| = \operatorname{dist}\left(z, \sigma(P_\Delta)\right)^{-1}$, see Proposition 2.32 in Chatelin (1983).

By Lemma 5.3 and the commuting between P_Δ and L we conclude

$$\begin{aligned}\|R(P_\Delta, z)f\|_{H^1} &\sim \|(\mathrm{Id} - L)^{1/2} R(P_\Delta, z) f\|_\mu \\ &\leq \|R(P_\Delta, z)\| \|(\mathrm{Id} - L)^{1/2} f\|_\mu \\ &\sim \operatorname{dist}(z, \sigma(P_\Delta))^{-1} \|f\|_{H^1}.\end{aligned}$$

Hence, $\|R(P_\Delta, z)\|_{H^1 \to H^1} \lesssim \rho^{-1}$ holds uniformly over $z \in B(\lambda, \rho)$ and $(\sigma, b) \in \Theta_s$. □

2.14. Remark. *The approximation error for the eigenvalues is much smaller. The Kato-Temple inequality (Theorem 6.21 in Chatelin (1983)) on $L^2(\mu)$ establishes the so-called super-convergence $|\lambda_1^J - \lambda_1| \lesssim 2^{-2Js}$ which is superoptimal for our purposes.*

2.4.4 Variance estimates

We have achieved the classical bias estimate in nonparametric estimation theory. In order to bound the stochastic error on the finite-dimensional space V_J, we return to vector-matrix notation and look for a bound on the error involved in the estimators $\hat{\mathbf{G}}$ and $\hat{\mathbf{P}}_\Delta$. The Euclidean norm is denoted by $\|\bullet\|_{l^2}$.

2.15. Lemma. *For any vector $\mathbf{v} \in \mathbb{R}^{|V_J|}$*

$$\mathbb{E}_{\sigma,b}[\|(\hat{\mathbf{G}} - \mathbf{G})\mathbf{v}\|_{l^2}^2] \lesssim \|\mathbf{v}\|_{l^2}^2 N^{-1} 2^J,$$

independently of $\mathbf{v}$ and uniformly over Θ_s.

Proof. We obtain by (5.1) in Lemma 5.2

$$\begin{aligned}
&\mathbb{E}_{\sigma,b}[\|(\hat{\mathbf{G}}-\mathbf{G})\mathbf{v}\|_{l^2}^2]\\
&=\sum_{|\lambda|\le J}\mathbb{E}_{\sigma,b}\Bigg[\bigg(\frac{1}{N}\Big(\tfrac{1}{2}\psi_\lambda(X_0)v(X_0)+\tfrac{1}{2}\psi_\lambda(X_{N\Delta})v(X_{N\Delta})+\sum_{n=1}^{N-1}\psi_\lambda(X_{n\Delta})v(X_{n\Delta})\Big)\\
&\quad-\mathbb{E}_{\sigma,b}[\psi_\lambda(X_0)v(X_0)]\bigg)^2\Bigg]\\
&\lesssim\sum_{|\lambda|\le J}N^{-1}\,\mathbb{E}_{\sigma,b}[(\psi_\lambda(X_0)v(X_0))^2]\\
&\lesssim N^{-1}\Big\|\sum_{|\lambda|\le J}\psi_\lambda^2\Big\|_\infty\|v\|_{L^2}^2\|\mu\|_\infty\\
&\lesssim N^{-1}2^J\|\mathbf{v}\|_{l^2}^2,
\end{aligned}$$

as asserted. □

2.16. Lemma. *For any vector* $\mathbf{v}$

$$\mathbb{E}_{\sigma,b}[\|(\hat{\mathbf{P}}_\Delta-\mathbf{P}_\Delta)\mathbf{v}\|_{l^2}^2]\lesssim\|\mathbf{v}\|_{l^2}^2N^{-1}2^J$$

holds, independently of $\mathbf{v}$ *and uniformly over* Θ_s.

Proof. We obtain by (5.2) in Lemma 5.2

$$\begin{aligned}
&\mathbb{E}_{\sigma,b}[\|(\hat{\mathbf{P}}_\Delta-\mathbf{P}_\Delta)v\|_{l^2}^2]\\
&=\sum_{|\lambda|\le J}\mathbb{E}_{\sigma,b}\Bigg[\bigg(\frac{1}{N}\sum_{n=1}^{N}\psi_\lambda(X_{(n-1)\Delta})v(X_{n\Delta})-\mathbb{E}_{\sigma,b}[\psi_\lambda(X_0)v(X_\Delta)]\bigg)^2\Bigg]\\
&\lesssim\sum_{|\lambda|\le J}N^{-1}\,\mathbb{E}_{\sigma,b}[(\psi_\lambda(X_0)v(X_1))^2]\\
&\le N^{-1}\sum_{|\lambda|\le J}\|\psi_\lambda\|_{L^2}^2\|v\|_{L^2}^2\|\mu p_\Delta\|_\infty\\
&\lesssim N^{-1}2^J\|\mathbf{v}\|_{l^2}^2,
\end{aligned}$$

as asserted. □

2.17. Definition. *We introduce the random set*

$$\mathcal{R}=\mathcal{R}_{J,N}:=\{\|\hat{\mathbf{G}}-\mathbf{G}\|\le\tfrac{1}{2}\|\mathbf{G}^{-1}\|^{-1}\}.$$

2.18. Remark. *Since* $\mathbf{G}$ *is invertible, so is* $\hat{\mathbf{G}}$ *on* $\mathcal{R}$ *with* $\|\hat{\mathbf{G}}^{-1}\|\le 2\|\mathbf{G}^{-1}\|$ *by the usual Neumann series argument.*

2.19. Lemma. *We find uniformly over* Θ_s

$$\mathbb{P}_{\sigma,b}(\Omega\setminus\mathcal{R})\lesssim N^{-1}2^{2J}.$$

Proof. By the classical Hilbert-Schmidt norm inequality,

$$\|\hat{\mathbf{G}} - \mathbf{G}\|^2_{l^2 \to l^2} \leq \sum_{|\lambda| \leq J} \|(\hat{\mathbf{G}} - \mathbf{G})\mathbf{e}_\lambda\|^2_{l^2 \to l^2}$$

holds with the unit vectors $(\mathbf{e}_\lambda)$ in $\mathbb{R}^{V_J}$. Then Lemma 2.15 gives the uniform moment estimate

$$\mathbb{E}_{\sigma,b}\big[\|\hat{\mathbf{G}} - \mathbf{G}\|^2_{l^2 \to l^2}\big] \lesssim N^{-1}2^{2J}.$$

Since the spaces $L^2([0,1])$ and $L^2(\mu)$ are isomorphic with uniform isomorphism constants, $\|\mathbf{G}^{-1}\| \sim 1$ holds uniformly over Θ_s and the assertion follows from Chebyshev's inequality. □

2.20. Proposition. *For any $\varepsilon > 0$ we have uniformly over Θ_s:*

$$\mathbb{P}_{\sigma,b}(\mathcal{R} \cap \{\|\hat{\mathbf{G}}^{-1}\hat{\mathbf{P}}_\Delta - \mathbf{G}^{-1}\mathbf{P}_\Delta\| \geq \varepsilon\}) \lesssim N^{-1}2^{2J}\varepsilon^{-2}.$$

Proof. First, we separate the different error terms:

$$\begin{aligned}\hat{\mathbf{G}}^{-1}\hat{\mathbf{P}}_\Delta - \mathbf{G}^{-1}\mathbf{P}_\Delta &= \hat{\mathbf{G}}^{-1}(\hat{\mathbf{P}}_\Delta - \mathbf{P}_\Delta) + (\hat{\mathbf{G}}^{-1} - \mathbf{G}^{-1})\mathbf{P}_\Delta \\ &= \hat{\mathbf{G}}^{-1}\big((\hat{\mathbf{P}}_\Delta - \mathbf{P}_\Delta) + (\mathbf{G} - \hat{\mathbf{G}})\mathbf{G}^{-1}\mathbf{P}_\Delta\big).\end{aligned}$$

On the set $\mathcal{R}$ we obtain by Remark 2.18

$$\begin{aligned}\|\hat{\mathbf{G}}^{-1}\hat{\mathbf{P}}_\Delta - \mathbf{G}^{-1}\mathbf{P}_\Delta\| &\leq \|\hat{\mathbf{G}}^{-1}\|\big(\|\hat{\mathbf{P}}_\Delta - \mathbf{P}_\Delta\| + \|\mathbf{G} - \hat{\mathbf{G}}\|\|\mathbf{G}^{-1}\|\|\mathbf{P}_\Delta\|\big) \\ &\leq 2\|\mathbf{G}^{-1}\|\big(\|\hat{\mathbf{P}}_\Delta - \mathbf{P}_\Delta\| + \|\mathbf{G} - \hat{\mathbf{G}}\|\|\mathbf{G}^{-1}\|\|\mathbf{P}_\Delta\|\big) \\ &\lesssim \|\hat{\mathbf{P}}_\Delta - \mathbf{P}_\Delta\| + \|\mathbf{G} - \hat{\mathbf{G}}\|.\end{aligned}$$

By Lemmas 2.15 and 2.16 and the Hilbert-Schmidt norm estimate (cf. the proof of Lemma 2.19) we obtain the uniform norm bound over Θ_s

$$\mathbb{E}_{\sigma,b}[\|\hat{\mathbf{G}}^{-1}\hat{\mathbf{P}}_\Delta - \mathbf{G}^{-1}\mathbf{P}_\Delta\|^2 \mathbf{1}_\mathcal{R}] \lesssim N^{-1}2^{2J}.$$

It remains to apply Chebyshev's inequality. □

Having established the weak consistency of the estimators in matrix norm, we now bound the error on the eigenspace.

2.21. Proposition. *Let $\mathbf{u}_1^J$ be the vector associated to the normalized eigenfunction u_1^J of $\pi_J^\mu P_\Delta$ with eigenvalue λ_1^J. Then uniformly over Θ_s the following risk bound holds*

$$\mathbb{E}_{\sigma,b}[\|(\hat{\mathbf{G}}^{-1}\hat{\mathbf{P}}_\Delta - \mathbf{G}^{-1}\mathbf{P}_\Delta^J)\mathbf{u}_1^J\|^2_{l^2} \mathbf{1}_\mathcal{R}] \lesssim N^{-1}2^J.$$

Proof. By the same separation of the error terms on $\mathcal{R}$ as in the preceding proof and by Lemmas 2.15 and 2.16 we find

$$\begin{aligned}&\mathbb{E}_{\sigma,b}[\|(\hat{\mathbf{G}}^{-1}\hat{\mathbf{P}}_\Delta - \mathbf{G}^{-1}\mathbf{P}_\Delta^J)\mathbf{u}_1^J\|^2_{l^2} \mathbf{1}_\mathcal{R}] \\ &\leq 8\|\mathbf{G}^{-1}\|^2\Big(\mathbb{E}\big[\|(\hat{\mathbf{P}}_\Delta - \mathbf{P}_\Delta^J)\mathbf{u}_1^J\|^2_{l^2}\big] + \mathbb{E}\big[\|(\mathbf{G} - \hat{\mathbf{G}})\lambda_1^J\mathbf{u}_1^J\|^2_{l^2}\big]\Big) \\ &\lesssim N^{-1}2^J.\end{aligned}$$

The uniformity over Θ_s follows from the respective statement in the lemmas. □

2.22. Corollary. *Let $\hat{\lambda}_1$ be the second largest eigenvalue of the matrix $\hat{\mathbf{G}}^{-1}\hat{\mathbf{P}}_\Delta$ with eigenvector $\hat{\mathbf{u}}_1$. If $\hat{\mathbf{G}}$ is not invertible or if $\|\hat{\mathbf{u}}_1\|_{l^2} \geq 2\sup_{\Theta_s}\|u_1\|_{L^2}$ holds, put $\hat{\mathbf{u}}_1 := \mathbf{0}$, $\hat{\lambda}_1 := 0$. If $N^{-1}2^{2J} \to 0$ holds, then uniformly over Θ_s the following bounds hold for $N, J \to \infty$:*

$$\mathbb{E}_{\sigma,b}\left[\left(|\hat{\lambda}_1 - \lambda_1^J|^2 + \|\hat{\mathbf{u}}_1 - \mathbf{u}_1^J\|_{l^2}^2\right)\mathbf{1}_{\mathcal{R}}\right] \lesssim N^{-1}2^J, \tag{2.18}$$

$$\mathbb{E}_{\sigma,b}\left[\|\hat{u}_1 - u_1^J\|_{H^1}^2\right] \lesssim N^{-1}2^{3J}. \tag{2.19}$$

Proof. For the proof of (2.18) we apply Proposition 2.9 using $\mathbb{R}^{V_J}$ with the Euclidean l^2-norm as Hilbert space and Corollary 2.10. Then Proposition 2.21 in connection with Proposition 2.20 (using $\varepsilon < R/2$ and $N^{-1}2^{2J} \to 0$) yields the correct asymptotic rate on the event $\mathcal{R}$. For the uniform choice of ρ and R for $\mathbf{G}^{-1}\mathbf{P}_\Delta^J$ in Proposition 2.9 just use the corresponding result for P_Δ and the convergence $\|\pi_J^\mu P_\Delta - P_\Delta\| \to 0$.

The precaution taken for undefined or too large $\hat{\mathbf{u}}_1$ is necessary for the event $\Omega \setminus \mathcal{R}$. Since the estimators $\hat{\lambda}_1$ and $\hat{\mathbf{u}}_1$ are now kept artificially bounded, the rate $\mathbb{P}_{\sigma,b}(\Omega \setminus \mathcal{R}) \lesssim N^{-1}2^{2J}$ established in Lemma 2.19 suffices to bound the risk on $\Omega \setminus \mathcal{R}$. Hence, the second estimate (2.19) is a consequence of (2.18) and the Bernstein inequality $\|\hat{u}_1 - u_1^J\|_{H^1} \lesssim 2^J\|\hat{u}_1 - u_1^J\|_{L^2}$ (see Appendix 5.2). □

2.23. Remark. *The main result of this section, namely* (2.19)*, can be extended to p-th moments for all $p \in (1,\infty)$:*

$$\mathbb{E}_{\sigma,b}\left[\|\hat{u}_1 - u_1^J\|_{H^1}^p\right]^{1/p} \lesssim N^{-1/2}2^{3J/2}. \tag{2.20}$$

Indeed, tracing back the steps, it suffices to obtain bounds on the moments of order p in Lemmas 2.15 and 2.16, which on their part rely on the mixing statement in Lemma 5.2. By arguments based on the Riesz convexity theorem this last lemma generalises to the corresponding bounds for pth moments of centred random variables H_1 and H_2, as derived in Section VII.4 of Rosenblatt (1971). It was only for the sake of clarity that we restricted to the case $p = 2$ here.

2.4.5 Proof of the upper bound for $\sigma(\bullet)$

Proof (Theorem 2.5). By Corollary 2.22 and our choice of J, $2^J \sim N^{1/(2s+3)}$,

$$\sup_{(\sigma,b)\in\Theta_s} \mathbb{E}_{\sigma,b}\left[|\hat{\lambda}_1 - \lambda_1^J|^2 + \|\hat{u}_1 - u_1^J\|_{H^1}^2\right] \lesssim N^{-1}2^{3J} \sim N^{-2s/(2s+3)}$$

holds. Using this estimate and the estimate for $\hat{\mu}$ in Proposition 2.8, the risk of the plug-in estimator $\widehat{\sigma^2(\bullet)}$ in (2.15) is bounded as asserted in the theorem. We only have to ensure that the stochastic error does not increase by the plug-in and that the denominator is uniformly bounded away from zero. Using the Cauchy-Schwarz inequality and Remark 2.23 on the higher moments of our estimators, we encounter no problem in the first case. The second issue is dealt with by

using the lower bound $c_{a,b} > 0$ in Proposition 5.5 so that an improvement of the estimate for the denominator by using

$$\widehat{\mu u}_1 := \max(\hat{\mu}\hat{u}_1, c_{a,b}) \tag{2.21}$$

instead of $\hat{\mu}\hat{u}_1$ guarantees the uniform lower bound away from zero. □

2.4.6 Upper bound for $b(\bullet)$

Since $b(\bullet) = S'(\bullet)/\mu(\bullet)$ holds, it suffices to discuss how to estimate $S'(\bullet)$, which amounts to estimating the eigenfunction u_1 in H^2-norm, compare with (2.7). Substituting H^2 for H^1 in Proposition 2.12 and its proof, we obtain the bound

$$\|\pi_J^\mu P_\Delta - P_\Delta\|_{H^2 \to H^2} \lesssim 2^{-J(s-1)},$$

because $\|\mathrm{Id} - \pi_J\|_{H^{s+1} \to H^2}$ is of this order. As in Corollary 2.13 this is also the rate for the bias estimate. The only fine point is the uniform norm equivalence $\|f\|_{H^2} \sim \|(\mathrm{Id} - L)f\|_\mu$ for $f \in \mathfrak{D}$, which follows by the methodology of perturbation and similarity arguments given in Section VI.4b of Engel and Nagel (2000). We omit further details.

The variance estimate is exactly the same. From (2.18) we infer by Bernstein's inequality for H^2 and the estimate of $\mathbb{P}_{\sigma,b}(\Omega \setminus \mathcal{R})$

$$\mathbb{E}_{\sigma,b}\left[\|\hat{u}_1 - u_1^J\|_{H^2}^2\right] \lesssim N^{-1} 2^{5J}.$$

Therefore balancing the bias and variance part of the risk by the choice $2^J \sim N^{1/(2s+3)}$ – as before – yields the asserted rate of convergence $N^{-(s-1)/(2s+3)}$.

3 Lower bounds

We prove an asymptotic lower bound for the estimation error of $\big(\sigma(\bullet), b(\bullet)\big)$. We keep up with the framework of Section 2. Let E_N denote the set of all estimators according to Definition 2.2.

3.1. Theorem. *The following asymptotic lower bound holds for all* $0 \le a < b \le 1$ *and* $s > 1$*:*

$$\liminf_{N\to\infty} N^{s/(2s+3)} \inf_{\widehat{\sigma^2} \in E_N} \sup_{(\sigma,b)\in\Theta_s} \mathbb{E}_{\sigma,b}\Big[\|\widehat{\sigma^2} - \sigma^2\|_{L^2([a,b])}^2\Big]^{1/2} > 0, \tag{3.1}$$

$$\liminf_{N\to\infty} N^{(s-1)/(2s+3)} \inf_{\hat{b} \in E_N} \sup_{(\sigma,b)\in\Theta_s} \mathbb{E}_{\sigma,b}\Big[\|\hat{b} - b\|_{L^2([a,b])}^2\Big]^{1/2} > 0. \tag{3.2}$$

Proof. The proof will be accomplished in several steps. Firstly, the usual Bayes prior technique is applied for the reduction to a problem of bounding certain likelihood ratios, see e.g. Korostelev and Tsybakov (1993). Then Proposition 5.4 is used so that only the L^2-distance between the transition probabilities remains to be estimated, which is finally done in the last step using Hilbert-Schmidt norm estimates and the explicit form of the inverse of the generator.

1. The idea is to perturb the drift and diffusion coefficients of the reflected Brownian motion in such a way that the invariant measure remains unchanged. Let us assume that ψ is a compactly supported wavelet in H^s with one vanishing moment. Without loss of generality we suppose $C > 1 > c > 0$ such that $(1,0) \in \Theta_s$ holds. We put $\psi_{jk} = 2^{j/2}\psi(2^j \bullet - k)$ and denote by $K_j \subset \mathbb{Z}$ a maximal set of indices k such that $\operatorname{supp}(\psi_{jk}) \subset [a,b]$ and $\operatorname{supp}(\psi_{jk}) \cap \operatorname{supp}(\psi_{jk'}) = \varnothing$ holds for all $k, k' \in K_j$, $k \neq k'$. Furthermore, we set $\gamma \sim 2^{-j(s+\frac{1}{2})}$ such that for all $\varepsilon = \varepsilon(j) \in \{-1,+1\}^{S_j}$

$$(\sqrt{2S_\varepsilon}, (S_\varepsilon)') \in \Theta_s \text{ with } S_\varepsilon(x) := S_\varepsilon(j,x) := \Big(2 + \gamma \sum_{k \in K_j} \varepsilon_k \psi_{jk}(x)\Big)^{-1}$$

holds. We consider the corresponding diffusions with generator

$$L_{S_\varepsilon} f(x) := (S_\varepsilon f')'(x) := S_\varepsilon(x) f''(x) + S'_\varepsilon(x) f'(x), \quad f \in \mathfrak{D}.$$

Hence, the invariant measure is the Lebesgue measure on $[0,1]$, in particular L_{S_ε} is selfadjoint on $L^2([0,1])$ for all ε.

The usual Assouad cube techniques, e.g. see Korostelev and Tsybakov (1993), give for any estimator $\hat{\sigma}(\bullet)$ and for $N \in \mathbb{N}$, $\rho > 0$, the lower bounds:

$$\sup_{(\sigma,b)\in\Theta_s} \mathbb{E}_{\sigma,b}\Big[\|\widehat{\sigma^2} - \sigma^2\|^2_{L^2([a,b])}\Big] \geq \frac{|K_j|}{2} \delta^2_{\sigma^2} e^{-\rho} p_0, \tag{3.3}$$

$$\sup_{(\sigma,b)\in\Theta_s} \mathbb{E}_{\sigma,b}\Big[\|\hat{b} - b\|^2_{L^2([a,b])}\Big] \geq \frac{|K_j|}{2} \delta^2_{b} e^{-\rho} p_0, \tag{3.4}$$

where the following definitions are used:

$$\begin{aligned}
&|K_j| := \text{cardinality of } K_j \sim 2^j, \\
&\delta_{\sigma^2} \leq \|2S_{\varepsilon'} - 2S_\varepsilon\|_{L^2} && \text{for all } \varepsilon, \varepsilon' \text{ with } \|\varepsilon - \varepsilon'\|_{l^1} = 2, \quad (3.5) \\
&\delta_b \leq \|S'_{\varepsilon'} - S'_\varepsilon\|_{L^2} && \text{for all } \varepsilon, \varepsilon' \text{ with } \|\varepsilon - \varepsilon'\|_{l^1} = 2, \quad (3.6) \\
&\mathbb{P}_\varepsilon := \mathbb{P}_{\sqrt{2S_\varepsilon}, S'_\varepsilon}, \\
&p_0 \leq \mathbb{P}_\varepsilon\Big(\tfrac{d\mathbb{P}_{\varepsilon'}}{d\mathbb{P}_\varepsilon}|_{\mathcal{F}_N} > e^{-\rho}\Big) && \text{for all } \varepsilon, \varepsilon' \text{ with } \|\varepsilon - \varepsilon'\|_{l^1} = 2. \quad (3.7)
\end{aligned}$$

In (3.5) we may choose asymptotically $\delta_{\sigma^2} \sim \gamma$ since for $x \in \operatorname{supp}(\psi_{jk})$ with $\varepsilon_k = -\varepsilon'_k$

$$S_{\varepsilon'}(x) - S_\varepsilon(x) = \pm 2\gamma \psi_{jk}(x) S_{\varepsilon'}(x) S_\varepsilon(x)$$

and $S_\varepsilon, S_{\varepsilon'} \to \frac{1}{2}$ holds uniformly so that the L^2-norm in (3.5) is indeed of order γ. In (3.6) we find equivalently $\delta_b \sim 2^j\gamma$. Due to $\gamma \sim 2^{-j(s+\frac{1}{2})}$, the proof of the theorem is accomplished once we have shown that in (3.7) a strictly positive p_0 can be chosen for fixed $\rho > 0$ and the asymptotics $2^j \sim N^{\frac{1}{2s+3}}$.

2. If we denote the transition probability densities $\mathbb{P}_\varepsilon(X_\Delta \in dy \,|\, X_0 = x)$ by $p_\varepsilon(x,y)dy$ and the transition density of reflected Brownian motion by p_{BM}, then we infer from Proposition 5.4

$$\lim_{j\to\infty} \sup_{\varepsilon\in\{-1,+1\}^{K_j}} \|p_\varepsilon - p_{BM}\|_\infty = 0$$

due to $\|S_\varepsilon - \frac{1}{2}\|_{C^1} \sim \gamma 2^{3j/2} \to 0$ for $s > 1$. We are now going to use the estimate $-\log(1+x) \le x^2 - x$, which is valid for all $x \ge -\frac{1}{2}$. For j so large that $\|1 - \frac{p_{\varepsilon'}}{p_\varepsilon}\|_\infty \le \frac{1}{2}$ holds, the Kullback-Leibler distance can be bounded from above (note that the invariant measure is Lebesgue measure):

$$\begin{aligned}
&\mathbb{E}_\varepsilon\Big[-\log\Big(\tfrac{d\,\mathbb{P}_{\varepsilon'}}{d\,\mathbb{P}_\varepsilon}|_{\mathcal{F}_N}\Big)\Big] \\
&= -\sum_{k=1}^{N} \mathbb{E}_\varepsilon\Big[\log\Big(\tfrac{p_{\varepsilon'}(X_{k-1},X_k)}{p_\varepsilon(X_{k-1},X_k)}\Big)\Big] \\
&= -N\int_0^1\int_0^1 \log\Big(\tfrac{p_{\varepsilon'}(x,y)}{p_\varepsilon(x,y)}\Big) p_\varepsilon(x,y)\,dy\,dx \\
&\le N\int_0^1\int_0^1 \frac{(p_{\varepsilon'}(x,y) - p_\varepsilon(x,y))^2}{p_\varepsilon(x,y)} - (p_{\varepsilon'}(x,y) - p_\varepsilon(x,y))\,dy\,dx \\
&= N\int_0^1\int_0^1 \frac{(p_{\varepsilon'}(x,y) - p_\varepsilon(x,y))^2}{p_\varepsilon(x,y)}\,dy\,dx \\
&\le N\|p_\varepsilon^{-1}\|_\infty \|p_{\varepsilon'} - p_\varepsilon\|^2_{L^2([0,1]^2)}.
\end{aligned}$$

The square root of the Kullback-Leibler distance bounds the total variation distance in order, see e.g. Deuschel and Stroock (1989), page 76, which by the Chebyshev inequality and the above estimate gives

$$\begin{aligned}
\mathbb{P}_\varepsilon\Big(\tfrac{d\,\mathbb{P}_{\varepsilon'}}{d\,\mathbb{P}_\varepsilon}|_{\mathcal{F}_N} > e^{-\rho}\Big) &= 1 - \mathbb{P}_\varepsilon\Big(\tfrac{d\,\mathbb{P}_{\varepsilon'}}{d\,\mathbb{P}_\varepsilon}|_{\mathcal{F}_N} - 1 \le e^{-\rho} - 1\Big) \\
&\ge 1 - \mathbb{E}_\varepsilon\Big[\Big|\tfrac{d\,\mathbb{P}_{\varepsilon'}}{d\,\mathbb{P}_\varepsilon}|_{\mathcal{F}_N} - 1\Big|\Big](1 - e^{-\rho})^{-1} \\
&= 1 - (1 - e^{-\rho})^{-1}\|(\mathbb{P}_{\varepsilon'} - \mathbb{P}_\varepsilon)|_{\mathcal{F}_N}\|_{TV} \\
&\ge 1 - CN^{1/2}\|p_{\varepsilon'} - p_\varepsilon\|_{L^2([0,1]^2)},
\end{aligned}$$

where $C > 0$ is some constant independent of γ, N, ε and j. Summarizing, we need the estimate

$$\limsup_{N,j\to\infty} N^{1/2}\|p_{\varepsilon'} - p_\varepsilon\|_{L^2([0,1]^2)} < C^{-1} \qquad \text{for } 2^j \sim N^{1/(2s+3)}.$$

Proposition 5.4 yields the rate $\gamma 2^{3j/2}$ for the supremum norm of the difference of the transition densities . However, this is far from optimal, because then the lower bound decays faster than $N^{-1/2}$.

3. By using genuine Hilbert space methods we shall establish a rate of order $\gamma 2^{-3j/2}$ for the L^2-distance of the transition densities. Observe first that $\|p_{\varepsilon'} - p_\varepsilon\|_{L^2([0,1]^2)}$ is exactly the Hilbert-Schmidt norm distance

$$\|P_\Delta^{\varepsilon'} - P_\Delta^\varepsilon\|_{HS}$$

between the transition operators derived from $L_{S_{\varepsilon'}}$ and L_{S_ε} acting on the Hilbert space $L^2([0,1])$. If we introduce

$$V := \Big\{ f \in L^2([0,1]) \,|\, \int f = 0 \Big\} \text{ and } V^\perp := \Big\{ f \in L^2([0,1]) \,|\, f \text{ constant} \Big\},$$

then the transition operators coincide on $V^\perp$ and leave the space V invariant so that $\|P_\Delta^{\varepsilon'} - P_\Delta^\varepsilon\|_{HS} = \|\big(P_\Delta^{\varepsilon'} - P_\Delta^\varepsilon\big)|_V\|_{HS}$.

We take advantage of the key result that for Lipschitz functions f with Lipschitz constant Λ on the union of the spectra of two selfadjoint bounded operators T_1 and T_2 the continuous functional calculus satisfies

$$\|f(T_1) - f(T_2)\|_{HS} \le \Lambda \|T_1 - T_2\|_{HS}, \tag{3.8}$$

see Kittaneh (1985). We proceed by bounding the Hilbert-Schmidt norm of the difference of the inverses of the generators and by then transferring this bound to the transition operators via (3.8). By the functional calculus for operators on V, the function $f(z) = \exp\big(\Delta(z^{-1})\big)$ sends $(L_\varepsilon|_V)^{-1}$ to $P_\Delta^\varepsilon|_V$. Moreover, f is uniformly Lipschitz continuous on $(-\infty, 0)$ due to $\Lambda := \sup_{z<0}|f'(z)| = 4\Delta^{-1}e^{-2} < \infty$. Thus, we arrive at

$$\|p_{\varepsilon'} - p_\varepsilon\|_{L^2([0,1]^2)} = \|\big(P_\Delta^{\varepsilon'} - P_\Delta^\varepsilon\big)|_V\|_{HS} \le \Lambda \|(L_{\varepsilon'}|_V)^{-1} - (L_\varepsilon|_V)^{-1}\|_{HS}.$$

The inverse of the generator L_ε on V has for $g \in V$ the explicit form

$$(L_\varepsilon|_V)^{-1} g(x) = \int_0^1 \left(\int_y^1 S_\varepsilon^{-1}(v)(v - \mathbf{1}_{[x,1]}(v))\, dv \right) g(y)\, dy. \tag{3.9}$$

Hence, using $|S_{\varepsilon'}^{-1} - S_\varepsilon^{-1}| = 2\gamma\psi_{jk}$ for some $k \in K_j$ and denoting by Ψ the primitive of ψ with compact support, we obtain

$$\begin{aligned}
&\|(L_{\varepsilon'}|_V)^{-1} - (L_\varepsilon|_V)^{-1}\|_{HS}^2 \\
&= \int_0^1 \int_0^1 \left(\int_y^1 2\gamma\psi_{jk}(v)(v - \mathbf{1}_{[x,1]}(v))\, dv \right)^2 dx\, dy \\
&= 4\gamma^2 2^{-j} \int_0^1 \int_0^1 \Big(-\Psi(2^j y - k)y - \int_y^1 \Psi(2^j v - k)\, dv + \\
&\qquad\qquad + \Psi(2^j(x \vee y) - k) \Big)^2 dx\, dy \\
&\lesssim \gamma^2 2^{-j} \|\Psi(2^j \bullet)\|_{L^2}^2 \sim \gamma^2 2^{-2j}.
\end{aligned}$$

Consequently, $\|p_{\varepsilon'} - p_\varepsilon\|_{L^2}^2 \sim 2^{-j(2s+3)}$ holds with an arbitrarily small constant if $\gamma 2^{j(s+\frac{1}{2})}$ is chosen sufficiently small. Hence, p_0 in (3.7) is strictly positive for this choice and the asymptotics $N 2^{-j(2s+3)} \to 1$, which remained to be proved.

□

4 Conclusion

In this short section, we address some questions that naturally arise from our study.

Kernel methods. The invariant density and the transition density could also be estimated using kernel methods, but the numerical calculation of the eigenpair $(\hat{\lambda}_1, \hat{u}_1)$ would then involve an additional discretisation step. Presumably, the risk can be kept to the same asymptotic order as for our projection approach.

Estimation at the boundary. Our plug-in estimator could only be defined on a proper subinterval of $(0,1)$. Estimation of $\sigma(\bullet)$ and $b(\bullet)$ at the boundary points leads to a risk increase, because the function $S(\bullet)$ satisfies

$$\lim_{x \downarrow 0} S^{-1}(x) = \frac{\nu_1 u_1(0) \mu(0)}{u_1''(0)}$$

by L'Hospital's rule applied to (2.5). Thus, the degree of ill-posedness for estimating $\sigma(0)$ and $b(0)$ is two and three, respectively, when using our plug-in estimators. On the other hand, a point-wise lower bound result – following the same lines as the L^2-lower bound proof – shows that this deterioration cannot be avoided.

Scalar diffusions over the real line. Can we extend our approach to scalar diffusions over the real line? In the case where Gaussian-type tail-estimates for the invariant density $\mu(\bullet)$ and polynomial growth bounds for the eigenfunction $u_1(\bullet)$ exist, we should get the same results in $L^2(\mu)$-loss, where $\mu(\bullet)$ is of course parameter-dependent. However, all the spectral approximation results shall be reconsidered with extra-care, in particular because of the possible absence of discrete spectrum for the generator $L = L_{\sigma,b}$. Some suitable results have been obtained by Chen and Wang (1997) and in the references given there.

Adaptation w.r.t. to unknown smoothness. Admittedly, the knowledge of the smoothness s that is needed for the construction of our estimators is unrealistic if practical purposes are considered. An adaptive estimation of the eigenpair $(u_1(\bullet), \lambda_1)$ and $\mu(\bullet)$ that yield adaptive estimators for $\big(\sigma(\bullet), b(\bullet)\big)$ could be obtained by the following modifications:

First, the adaptive estimation of $\mu(\bullet)$ in a classical mixing framework is fairly well known (see for instance Viennet (1997) by model selection, or Tribouley

and Viennet (1998) by wavelet thresholding). Second, taking advantage of the multiresolution structure provided by wavelets, the adaptive estimation of P_Δ could be obtained by introducing an appropriate thresholding in the estimated matrices on a large approximation space. As already mentioned, a model selection approach is natural as well. Presumably, the methods developed in Baraud (2000) or Baraud, Comte and Viennet (2001) would provide us with an adaptive estimator $\hat{P}_\Delta$ of P_Δ. All further steps needed to derive $\widehat{\sigma^2}(\bullet)$ and $\hat{b}(\bullet)$ are analytic and do not involve any further estimation so that these estimators will be adaptive, too.

Estimating $b(\bullet)$ when $\sigma(\bullet)$ is known and vice versa. It is noteworthy that sin the continuous time or high frequency observation case, the parameter $b(\bullet)$ does not influence the asymptotic behaviour of the estimator of $\sigma(\bullet)$ and vice versa. The estimation problems are separated. In particular, the knowledge of the other parameter does not yield better minimax rates of convergence. In our low frequency regime we had to suppose tight regularity connections between $\sigma(\bullet)$ and $b(\bullet)$ in the upper as well as the lower bound proof. This stems from the fact that for the underlying Markov chain $X_0, X_\Delta, \ldots, X_{N\Delta}$ the parameters $\mu(\bullet)$ and $S(\bullet)$ are more natural and the regularity of these functions depends on both, the regularity of $b(\bullet)$ and of $\sigma(\bullet)$.

What happens in the low frequency regime if we know one of the parameters? If σ is known, we only need an estimate $\hat{\mu}$ of the invariant density, since

$$b(x) = \frac{\left(\sigma^2(x)\mu(x)\right)'}{2\mu(x)}, \quad x \in [0,1].$$

Estimation of $\mu \in H^s$, $s > 1$, in H^1-norm can be achieved with rate $N^{-(s-1)/(2s+1)}$ due to the degree one of ill-posedness involved. By the identity just given, this rate is thus also valid for estimating $b(\bullet)$ in L^2-norm, assuming σ to be known.

For known drift coefficient $b(\bullet)$ the estimation of $\sigma(\bullet)$ is a little bit more complex. We know

$$\sigma^2(x) = 2\frac{\int_0^x b(y)\mu(y)\,dy + C}{\mu(x)}, \quad x \in [0,1],$$

where C is a suitable constant. If we knew $C = \sigma^2(0)$, we would obtain the classical rate $N^{-s/(2s+1)}$ provided $\mu \in H^s$. In order to estimate the one-dimensional parameter C inference based on the transition operator P_Δ has to be drawn. Using a preliminary nonparametric estimate $\widehat{\sigma^2}_C$ depending on the parameter C and then fitting a parametric model for C in the spirit of Kessler and Sørensen, we are likely to find the same rate for unknown C.

In any case, note that the model assumption of knowing one parameter exactly is highly artificial so that proper proofs of minimax rates do not seem to be worthwhile.

5 Appendix

5.1 Scalar diffusions

We shall need several technical results, mainly to describe the dependence of certain quantities on the underlying diffusion parameters. The following result is in close analogy with Section IV.5 in Bass (1998).

5.1. Lemma. *The second largest eigenvalue ν_1 of the infinitesimal generator $L_{\sigma,b}$ can be bounded away from zero:*

$$\nu_1 \leq - \inf_{x\in[0,1]} S(x) =: -s_0.$$

This eigenvalue is simple and the corresponding eigenfunction f_1 is monotone.

Proof. The variational characterisation of ν_1 (Davies (1996), Section 4.5) and partial integration yield:

$$\nu_1 = \sup_{\substack{\|f\|_\mu=1\\ \langle f,\mathbf{1}\rangle_\mu=0}} \langle Lf, f\rangle_\mu = - \inf_{\substack{\|f\|_\mu=1\\ \langle f,\mathbf{1}\rangle_\mu=0}} \int_0^1 S(x) f'(x)^2\, dx.$$

Given the derivative f', the function $f \in \mathfrak{D}$ with $\langle f, \mathbf{1}\rangle_\mu = 0$ is uniquely determined. Setting $M(x) := \mu([0,x])$, this function f satisfies

$$0 = \int_0^1 \left(f(0) + \int_0^x f'(y)\, dy \right) \mu(x)\, dx = f(0) - \int_0^1 f'(y)(1 - M(y))\, dy.$$

For two functions f, g in the $L^2(\mu)$-orthogonal space of the constants we therefore find

$$\begin{aligned}
\langle f, g\rangle_\mu &= \int_0^1 \left(f(0) + \int_0^x f'(y)\, dy \right) \left(g(0) + \int_0^x g'(z)\, dz \right) \mu(x)\, dx \\
&= f(0)g(0) + f(0) \int_0^1 g'(z)(1 - M(z))\, dz + g(0) \int_0^1 f'(y)(1 - M(y))\, dy \\
&\quad + \int_0^1 \int_0^1 f'(y)g'(y)(1 - M(y \wedge z))\, dy\, dz \\
&= \int_0^1 \int_0^1 (M(y \wedge z) - M(y)M(z))\, f'(y)g'(z)\, dy\, dz \\
&=: \int_0^1 \int_0^1 m(y,z) f'(y)g'(z)\, dy\, dz.
\end{aligned}$$

The kernel $m(y,z)$ is positive on $(0,1)^2$ and bounded by 1 whence we obtain by regarding $u = f'$

$$-\nu_1 = \inf_u \frac{\int_0^1 S(x)u(x)^2\, dx}{\int_0^1 \int_0^1 m(y,z)u(y)u(z)\, dy\, dz} \geq \frac{s_0 \|u\|_{L^2}^2}{\|u\|_{L^1}^2} \geq s_0.$$

If the derivative of the eigenfunction f_1' changed sign, we could write $f_1' = u^+ - u^-$ with two nonnegative functions u^+, u^- that are nontrivial. However, this would entail that the antiderivative f_0 of $f_0' := u^+ + u^-$ satisfies $\langle Lf_0, f_0\rangle_\mu = \langle Lf_1, f_1\rangle_\mu$, while $\|f_0\|_\mu$ would be strictly greater than $\|f_1\|_\mu$ due to the positivity of $m(u,v)$. This is in contradiction with the variational characterisation of ν_1 so that $f_1' \geq 0$ or $f_1' \leq 0$ has to hold on the whole interval $[0,1]$. In particular, all eigenfunctions corresponding to ν_1 are monotone which shows that for any two eigenfunctions f_1 and g_1 the integrand in

$$\langle f_1, g_1\rangle_\mu = \int_0^1 \int_0^1 m(y,z) f_1'(y) g_1'(z)\, dy\, dz$$

does not change sign and the whole integral does not vanish. We infer that the eigenspace of ν_1 cannot contain two orthogonal functions and is thus one-dimensional. □

5.2. Lemma. *For H_1, $H_2 \in L^2([0,1])$ we have the following two uniform variance estimates over Θ_s:*

$$\mathrm{Var}_{\sigma,b}\Big[\frac{1}{N}\sum_{n=1}^N H_1(X_{n\Delta})\Big] \lesssim N^{-1}\,\mathbb{E}_{\sigma,b}[H_1(X_0)^2], \tag{5.1}$$

$$\mathrm{Var}_{\sigma,b}\Big[\frac{1}{N}\sum_{n=1}^N H_1(X_{(n-1)\Delta})H_2(X_{n\Delta})\Big] \lesssim N^{-1}\,\mathbb{E}_{\sigma,b}[H_1(X_0)^2 H_2(X_1)^2]. \tag{5.2}$$

Proof. Due to the uniform spectral gap s_0 over Θ_s (Proposition 5.1), the transition operator P_Δ satisfies $\|P_\Delta f\|_\mu \leq \gamma \|f\|_\mu$ with $\gamma := e^{-|s_0|\Delta} < 1$ for all f with $\langle f, \mathbf{1}\rangle_\mu = 0$.

We obtain by regarding the centred random variable $f_1 := H_1 - \mathbb{E}_{\sigma,b}[H_1]$

$$\begin{aligned}
\mathrm{Var}_{\sigma,b}\Big[\sum_{n=1}^N H_1(X_{n\Delta})\Big] &= \sum_{m,n=1}^N \mathbb{E}_{\sigma,b}[f_1(X_{m\Delta}) f_1(X_{n\Delta})] \\
&= \sum_{m,n=1}^N \langle f_1, P_\Delta^{|m-n|} f_1\rangle_\mu \\
&\leq \sum_{m,n=1}^N \gamma^{|m-n|} \|f_1\|_\mu^2 \\
&\lesssim N\,\mathbb{E}_{\sigma,b}[H_1(X_0)^2],
\end{aligned}$$

which after dividing by N^2 is the first estimate.

The second estimate follows along the same lines. Merely observe that for $m > n$ by the projection property of conditional expectations

$$\begin{aligned}
&\mathbb{E}_{\sigma,b}[H_1(X_{(n-1)\Delta})H_2(X_{n\Delta})H_1(X_{(m-1)\Delta})H_2(X_{m\Delta})] = \\
&\qquad\qquad \langle H_1 \bullet (P_\Delta H_2), P_\Delta^{m-n-1}(H_1 \bullet (P_\Delta H_2))\rangle_\mu
\end{aligned}$$

holds where $\bullet$ is the usual multiplication operator. □

5.3. Lemma. *Uniformly over Θ_s the following norm equivalence holds true:*

$$\|f\|_{H^1} \sim \|(\mathrm{Id} - L)^{-1/2} f\|_\mu \qquad \text{for all } f \in H^1.$$

Proof. The invariant measure μ and the function S are uniformly bounded away from zero and infinity so that we obtain with uniform constants for $f \in \mathfrak{D}$

$$\begin{aligned}
\|f\|_{H^1}^2 &= \langle f, f\rangle + \langle f', f'\rangle \\
&\sim \langle f, f\rangle_\mu + \langle Sf', f'\rangle \\
&= \langle f, f\rangle_\mu - \langle (Sf')', f\rangle \\
&= \langle (\mathrm{Id} - L) f, f\rangle_\mu \\
&= \|(\mathrm{Id} - L)^{1/2} f\|_\mu^2.
\end{aligned}$$

By an approximation argument this equivalence extends to all $f \in H^1 = \mathrm{dom}(L^{1/2})$. □

The next proposition is well known for parabolic equations on the real axis (Theorem 15 in Friedman (1964)), but seems to require a proof for equations with Neumann boundary conditions.

5.4. Proposition. *Suppose* $\big((\sigma_n(\bullet), b_n(\bullet)\big) \in \Theta_s$, $n \geq 0$, *and*

$$\lim_{n\to\infty} \|\sigma_n - \sigma_0\|_\infty = 0, \quad \lim_{n\to\infty} \|b_n - b_0\|_\infty = 0.$$

denote the corresponding transition densities by $(p_t^{(n)})$. *Then these transition probabilities converge uniformly*

$$\lim_{n\to\infty} \|p_t^{(n)}(\bullet, \bullet) - p_t^{(0)}(\bullet, \bullet)\| = 0.$$

Proof. From the definitions (2.2) and (2.3) of invariant density $\mu(\bullet)$ and function $S(\bullet)$ we infer the convergences $\|\mu_n - \mu_0\|_\infty \to 0$ and $\|S_n - S_0\|_{C^1} \to 0$ with obvious notation. By Theorem VI.4.6 in Engel and Nagel (2000), the semigroups considered on $C([0,1])$ are all analytic. Thus, the difference of the generators $L^{(n)} - L^{(0)}$ is $L^{(0)}$-bounded, since for large n and $f \in \mathrm{ran}(L^{(0)})$:

$$\begin{aligned}
&\|(L^{(n)}(L^{(0)})^{-1} f - f\|_\infty \\
&= \left\| L^{(n)} \Big(C + \int_0^x S_0^{-1}(y) \int_0^y \mu_0(z) f(z)\, dz\, dy \Big) - f \right\|_\infty \\
&= \left\| \mu_n^{-1}(x) \Big(S_n(x) S_0^{-1}(x) \int_0^x \mu_0(z) f(z)\, dz \Big)' - f(x) \right\|_\infty \\
&= \left\| \mu_n^{-1}(x) \Big(S_n(x) S_0^{-1}(x) \mu_0(x) f(x) + (S_n S_0^{-1})'(x) \int_0^x \mu_0(z) f(z)\, dz \Big) - f(x) \right\|_\infty \\
&\leq \left(\left\| \tfrac{\mu_0 S_0^{-1} - \mu_n S_n^{-1}}{\mu_n S_n^{-1}} \right\|_\infty + \left\| \tfrac{S_n' S_0 - S_n S_0'}{S_0^2} \right\|_\infty \right) \|f\|_\infty \\
&=: \varepsilon_n \|f\|_\infty,
\end{aligned}$$

which implies $\|(L^{(n)} - L^{(0)})g\| \leq \varepsilon_n \|L^{(0)} g\| \to 0$ (put $g = (L^{(0)})^{-1} f$). The proof of Kato (1995), Theorem IX.2.4, then shows the existence of a constant $M > 0$ such that

$$\|(L^{(n)} - \zeta)^{-1}\| \leq \frac{M}{\zeta(1 - \varepsilon_n(1+M))}, \quad \zeta \in \mathbb{C},\ |\arg(\zeta)| \leq \tfrac{3}{4}\pi; \quad n \in \mathbb{N}. \quad (5.3)$$

The Neumann series and (5.3) give for test functions f

$$\begin{aligned}
\|((L^{(n)} - \zeta)^{-1} - (L^{(0)} - \zeta)^{-1}) f\|_\infty &\leq \sum_{k=1}^{\infty} (\varepsilon_n)^k \|(L^{(0)} - \zeta)^{-1}\|^k \|(L^{(0)} - \zeta)^{-1} f\|_{L^1} \\
&= \frac{\varepsilon_n \|(L^{(0)} - \zeta)^{-1}\|}{1 - \varepsilon_n \|(L^{(0)} - \zeta)^{-1}\|} \|(L^{(0)} - \zeta)^{-1} f\|_\infty \\
&\leq \frac{\varepsilon_n M}{\zeta - \varepsilon_n M} \|(L^{(0)} - \zeta)^{-1} f\|_\infty .
\end{aligned}$$

By Lemma 5.3 and duality of (H^1, H^{-1}) we find

$$\begin{aligned}
\|(L^{(0)} - \zeta)^{-1} f\|_\infty &\leq \|(L^{(0)} - \zeta)^{-1} f\|_{H^1} \\
&\sim \|(1 - L^{(0)})^{1/2} (L^{(0)} - \zeta)^{-1} f\|_\mu \\
&= \|(L^{(0)} - 1)(L^{(0)} - \zeta)^{-1}\|_{L^2(\mu) \to L^2(\mu)} \|(1 - L^{(0)})^{-1/2} f\|_\mu \\
&\lesssim \zeta^{-1} \|f\|_{H^{-1}} .
\end{aligned}$$

This implies by the integral representation of analytic semigroups (Eq. IX.1.50 in Kato (1995)) and the dominated convergence theorem the norm convergence of the semigroup operators as ε_n tends to zero:

$$\lim_{n \to \infty} \|P_t - P_t^{(n)} f\|_\infty \lesssim \varepsilon_n \|f\|_{H^{-1}} \to 0,$$

uniformly for t on compact intervals. Consequently, approximating the Dirac measure $\delta_y \in H^{-1}$ at $y \in [0,1]$ by H^{-1}-bounded test functions, we infer that the kernels converge uniformly:

$$\lim_{n \to \infty} \operatorname{esssup}_{x,y} |p_t(x,y) - p_t^{(n)}(x,y)| = 0.$$

□

However, this proposition is not optimal in the sense that even the convergence in weaker functional or distributional norms of the drift and diffusion coefficients suffice to obtain the convergence of the semigroups. Of course, this kind of result is not surprising. In the case of the domain $\mathbb{R}^d$, Bass (1998) in Chapter VI investigates the weak convergence of the diffusion laws for pointwise convergence of the drift and diffusion coefficients.

5.5. Proposition. *For the class Θ_s there is a constant $\rho > 0$ such that for all parameters $(\sigma(\bullet), b(\bullet))$ the eigenvalue $\lambda_1 = \lambda_1(\sigma, b)$ of P_Δ is uniformly separated:*

$$\lambda_1 + 2\rho \leq 1, \qquad \lambda_1 - 2\rho \geq \lambda_2.$$

Furthermore, for all $0 < a < b < 1$ there is a uniform constant $c_{a,b} > 0$ such that the associated first eigenfunction $u_1 = u_1(\sigma, b)$ satisfies for all $(\sigma, b) \in \Theta_s$

$$\min_{x \in [a,b]} |u_1'(x)| \geq c_{a,b}.$$

Proof. Proceeding indirectly, assume that there is a sequence $(\sigma_n, b_n) \in \Theta_s$ such that the corresponding eigenvalues satisfy $\lambda_1^{(n)} \to 1$ (or $\lambda_1^{(n)} - \lambda_2^{(n)} \to 0$, respectively). By the compactness of the Sobolev embedding of H^s into C^0 we can pass to a uniformly converging subsequence. Hence, Proposition 5.4 yields that the corresponding transition densities converge uniformly, which implies that the transition operators $P_\Delta^{(n)}$ converge in operator norm on $L^2([0,1])$. By Proposition 5.6 and Theorem 5.20 in Chatelin (1983), this entails the convergence of their eigenvalues with preservation of the multiplicities. Since the limiting operator is again associated with an elliptic reflected diffusion, the fact that the eigenvalue $\lambda_1 = e^{\Delta \nu_1}$ is always simple (Lemma 5.1) gives the desired contradiction.

By the same indirect arguments, we construct transition operators $P_\Delta^{(n)}$ on the space $\mathcal{C}([0,1])$ and infer that the eigenfunctions $u_1^{(n)}$ (Theorem 5.10 in Chatelin (1983)), the invariant measures $\mu^{(n)}$ (see (2.2)) and the inverses of the functions $S^{(n)}$ (see (2.3) converge in supremum norm. Therefore $(u_1^{(n)})' = \nu_1^{(n)} (S^{(n)})^{-1} \int u_1^{(n)} \mu^{(n)}$ also converges in supremum norm. Due to $u_1'|_{[a,b]} \neq 0$ (Lemma 5.1) this implies that $(u_1^{(n)})'$ cannot converge to zero on $[a,b]$. □

5.6. Lemma. *The $L^2(\mu)$-normalized eigenfunction u_k of L corresponding to the $(k+1)$st largest eigenvalue ν_k satisfies*

$$\|u_k\|_{H^{s+1}} \leq C(s, s_0, \|S^{-1}\|_s, \|\mu\|_{s-1}) |\nu_k|^{\lceil s \rceil},$$

where C is a continuous function of its arguments.

Proof. We know that $\mu^{-1}(S u_k')' = \nu_k u_k$ and $u_k'(0) = 0$ holds, which implies

$$u_k'(x) = \nu_k S^{-1}(x) \int_0^x u_k(u) \mu(u)\, du.$$

suppose $u_k \in H^{r+1}$ with $r \in [0, s]$. Then the function $u_k \mu$ is in $H^{r \wedge (s-1)}$ due to $u_k \in C^r$ (Sobolev embeddings). Hence, the antiderivative is in $H^{(r+1) \wedge s}$. As $S^{-1} \in H^s$ holds, the right hand side is an element of $H^{s \wedge (r+1)}$. We conclude that the regularity r of u_k is by one larger which implies that u_k is in H^{s+1}.

In quantitative terms we obtain for $r \in [1,s]$, where we use the seminorm $|f|_s := \|f^{(s)}\|_{L^2}$:

$$\begin{aligned}
|u_k'|_r &\le |\nu_k| C(r) \left(|S^{-1}|_r \left\| \int_0^\bullet u_k \mu \right\|_\infty + \|S^{-1}\|_\infty \left| \int_0^\bullet u_k \mu \right|_r \right) \\
&\le |\nu_k| C(r) \|S^{-1}\|_r (\|u_k\|_{L^1(\mu)} + |u_k \mu|_{r-1}) \\
&\le |\nu_k| C(r) \|S^{-1}\|_s (1 + |u_k|_{r-1} \|\mu\|_\infty + \|u_k\|_\infty \|\mu\|_{r-1} \\
&\le |\nu_k| C(r) \|S^{-1}\|_s (1 + 2\|u_k\|_{r-1} \|\mu\|_{s-1}).
\end{aligned}$$

By applying this estimate $\|u_k\|_{r+1} \lesssim |\nu_k| \|S^{-1}\|_s (1 + \|u_k\|_{r-1} \|\mu\|_{s-1})$ successively for $r = 1, 2, \dots, \lfloor s \rfloor$ and finally for $r = s-1$, the estimate follows from the normalisation $\|u_k\|_0 = 1$. □

5.7. Proposition. *For $(\sigma, b) \in \Theta_s$ the corresponding transition probability density $p_\Delta = p_{\Delta,\sigma,b}$ satisfies*

$$\sup_{(\sigma,b)\in\Theta_s} \|p_\Delta\|_{H^{s+1} \times H^s} < \infty.$$

Proof. The spectral decomposition of $P_\Delta : L^2(\mu) \to L^2(\mu)$ yields the kernel representation

$$p_\Delta(x,y) = \mu(y) \sum_{k=0}^\infty e^{\nu_k \Delta} u_k(x) u_k(y), \quad x, y \in [0,1].$$

Due to the uniform ellipticity and uniform boundedness of P_Δ we know that $\nu_k \in [-C_1 k^2, -C_2 k^2]$ with uniform constant $C_1, C_2 > 0$ on Θ_s. From the preceding Lemma 5.6 and the Sobolev embedding $H^{s+1} \subset C^s$ we infer

$$\begin{aligned}
\|p_\Delta\|_{H^{s+1} \times H^s} &\le \sum_{k=0}^\infty e^{-C_2 \Delta k^2} \|u_k\|_{s+1} \|\mu u_k\|_s \\
&\le \sum_{k=0}^\infty C(s, s_0, \|S^{-1}\|_s, \|\mu\|_{s-1})^2 e^{-C_2 \Delta k^2} (C_1 k^2)^{s+1} \|\mu\|_s,
\end{aligned}$$

which gives the desired uniform estimate. □

5.2 Wavelets, Sobolev spaces and approximation results

For the whole topic we refer to the monograph by Cohen (2000). Let us introduce the scale of Sobolev spaces $H^s(\mathcal{I})$, $s \ge 0$, $\mathcal{I} \subset \mathbb{R}$ an interval, which combine the advantages of a differentiability description and of a Hilbert space structure for function classes. For $s = m \in \mathbb{N}_0$ set

$$H^m(\mathcal{I}) := \{f \in L^2(\mathcal{I}) \mid f^{(i)} \in L^2(\mathcal{I}) \text{ for all } i = 0, \dots, m\},$$

where $f^{(i)}$ denotes the i-th derivative of f in a weak (distributional) sense. These spaces are Hilbert spaces with respect to the following norm and scalar product

$$\|f\|_m^2 := \sum_{i=0}^{m} \|f^{(i)}\|_{L^2}^2, \quad \langle f, g\rangle_m := \sum_{i=0}^{m} \langle f^{(i)}, g^{(i)}\rangle_{L^2}.$$

The abstract complex interpolation method yields Sobolev spaces $H^s(\mathcal{I})$ of all real orders $s \geq 0$. We start with the definition of an orthonormal wavelet basis in $L^2(\mathbb{R})$.

5.8. Definition. *For $j \in \mathbb{Z}$ and $k \in \mathbb{Z}$ introduce the multi-index $\lambda = (j,k)$ and put $|\lambda| := |(j,k)| := j$. A* wavelet basis $(\psi_\lambda)_\lambda$ *is an orthonormal basis of functions in $L^2(\mathbb{R})$, derived from one function $\psi \in L^2(\mathbb{R})$ by translations and dilations*

$$\psi_\lambda(x) := \psi_{jk}(x) := 2^{j/2}\psi(2^j x - k).$$

The existence of wavelet bases besides the Haar wavelet and related spline wavelets is a nontrivial fact, in particular compactly supported wavelets of arbitrary high regularity exist, the so-called Daubechies wavelets. Orthonormal wavelet bases on a bounded interval $\mathcal{I}$ are constructed similarly. The basis functions are obtained by restricting the Daubechies wavelets to this interval. Wavelet functions ψ_λ whose support crosses the boundary of $\mathcal{I}$ are suitably corrected in order to keep the orthogonality and approximation properties. These corrected functions are still denoted by ψ_λ even if they are not directly derived from ψ. A consequence of this construction is that only multi-indices $\lambda = (j,k)$ with $|k| \lesssim 2^j$ are used and that the approximation spaces

$$V_j := \operatorname{span}\{\psi_\lambda \,|\, |\lambda| \leq j\}$$

are finite-dimensional, whence we can start off with a space V_{-1}, which always contains the indicator function $\mathbf{1}_\mathcal{I}$. Then any function $f \in L^2(\mathcal{I})$ has the wavelet decomposition

$$f = \sum_\lambda \langle f, \psi_\lambda\rangle \psi_\lambda = \sum_{j \geq -1} \sum_k \langle f, \psi_{jk}\rangle \psi_{jk},$$

where the second sum is taken over all k with $\operatorname{supp}(\psi_{jk}) \cap \mathcal{I} \neq \varnothing$. Note that summation over $|\lambda| \leq j_0$ will always mean summation over (j,k) for all $j \leq j_0$ and all corresponding values of k.

Wavelets exactly describe functions in the L^2-Sobolev spaces H^s for any real $s \geq 0$. Denote by π_j the L^2-orthogonal projection onto V_j.

5.9. Definition. *A wavelet basis (ψ_λ) will be called* s-regular *on the interval $\mathcal{I}$ if the following two conditions are satisfied:*

1. *The function f is an element of $H^s(\mathcal{I})$ if and only if*

$$\left(\|\pi_{-1} f\|_{L^2}^2 + \sum_{j=0}^{\infty} 2^{2sj} \sum_k |\langle f, \psi_{jk}\rangle|^2\right)^{1/2} < \infty,$$

and this expression constitutes a norm equivalent to the H^s-norm.

2. For all $k = 0, \dots, \lfloor s \rfloor$ the vanishing moment property is fulfilled

$$\int_{-\infty}^{\infty} x^k \psi(x)\, dx = 0.$$

From Cohen (2000) we know (only mind the slightly different notion of s-regularity there):

5.10. Theorem. *s-regular wavelet bases exist for any $s > 0$. Moreover, they may be chosen to have compact support.*

The multiresolution spaces V_n spanned by s-regular wavelets satisfy the Jackson and Bernstein inequalities, also called direct and inverse estimates:

$$\|(\mathrm{Id} - \pi_J) f\|_{H^t} \lesssim 2^{-J(s-t)} \|f\|_{H^s}, \quad 0 \le t \le s, \tag{5.4}$$

$$\forall\, v_J \in V_J : \ \|v_J\|_{H^s} \lesssim 2^{J(s-t)} \|v_J\|_{H^t}, \quad 0 \le t \le s. \tag{5.5}$$

The proof of these inequalities is immediate from the property of s-regularity, which by interpolation implies t-regularity for all $0 \le t \le s$.

References

[1] Banon, G. (1978) Nonparametric identification for diffusion processes. *SIAM Journal of Control and Optimization*, 16, 380-395.

[2] Baraud, Y. (2000) Model selection for regression on a fixed design. *Probab. Theory Related Fields*, 117, 467-493.

[3] Baraud, Y., Comte, F. and Viennet, G. (2001) Adaptive estimation in autoregression or β-mixing regression via model selection. *Ann. Statist.* 29, 839–875.

[4] Bass, R. F., (1998) *Diffusions and elliptic operators.* Probability and its Applications. Springer, New York.

[5] Birgé, L. (1983) Approximation dans les espaces métriques et théorie de l'estimation. *Z. Warsch. Ver. Geb.*, 65, 181-237.

[6] Brown, B.M. and Hewitt, J.I. (1975) Asymptotic likelihood theory for diffusion processes. *J. Appl. Prob.*, 12, 228-238.

[7] Cepa, E. (1995) *Équations différentielles stochastiques multivoques. (French) [Multivariate stochastic differential equations]* Séminaire de Probabilités, XXIX, 86-107, Lecture Notes in Math., 1613, Springer, Berlin.

[8] Chatelin, F. (1983) *Spectral approximation of linear operators.* Computer Science and Applied Mathematics. Academic Press, New-York - London.

[9] Chen, M. and Wang, F. (1997) *Estimation of spectral gap for elliptic operators*, Transactions of the AMS, 349(3), 1239-1267.

[10] Chen, X. (1998) Lecture given at the Paris-Berlin conference on financial mathematics and adaptive estimation, September 1998, Garchy, France.

[11] Cohen, A. (2000) *Wavelet methods in numerical analysis.* In: *Handbook of numerical analysis*, Vol. 7, ed. by P.G. Ciarlet, North Holland/Elsevier, Amsterdam.

[12] Dacunha-Castelle, D. and Florens, D. (1986) Estimation of the coefficients of a diffusion from discrete observations. *Stochastics*, 19, 263-284.

[13] Dalalyan, A. (2001) *Estimation asymptotiquement efficace pour des processus de diffusion ergodique (in English).* PhD Thesis, Université du Maine, France.

[14] Davies, E. B. (1996) *Spectral theory and differential operators.* Cambridge Studies in Advanced Mathematics. 42. Cambridge Univ. Press, Cambridge.

[15] Deuschel, J.D. and Stroock, W. (1989) *Large deviations.* Academic Press, London.

[16] Ethier, S. N. and Kurtz T.G. (1986) *Markov processes. Characterization and convergence.* Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, New-York.

[17] Engel, K-J. and Nagel, R. (2000) *One-parameter semigroups for linear evolution equations.* Graduate texts in Mathematics. 194. Springer, Berlin.

[18] Friedman, A. (1964) *Partial differential equations of parabolic type.* Prentice-Hall, New Jersey.

[19] Gobet, E. (2002) LAN property for ergodic diffusions with discrete observations. Forthcoming in *Annales de l'IHP – Probabilités et Statistique.*

[20] Hansen, L.P. and Scheinkman, J.A. (1995) Back to the future: Generating moment implications for continuous time markov processes. *Econometrica*, 63 (4), 767-804.

[21] Hansen, L.P., Scheinkman, J.A. and Touzi, N. (1998) Spectral methods for identifying scalar diffusions. *J. of Econometrics*, 86, 1-32.

[22] Hoffmann, M. (1999) Adaptive estimation in diffusion processes. *Stoch. Proc. Appl.*, 79, 135-163.

[23] Ibragimov, I.A. and Khas'minskii, R.Z. (1981) *Statistical Estimation, Asymptotic theory.* Springer, Berlin.

[24] Kato, T. (1995) *Perturbation theory for linear operators. Reprint of the corr. print. of the 2nd ed. 1980.* Classics in Mathematics. Springer-Verlag, Berlin.

[25] Kessler, M. and Sørensen, M. (1999) Estimating equations based on eigenfunctions for a discretely observed diffusion process. *Bernoulli*, 5, 299-314.

[26] Kessler, M. (1997). Estimation of an ergodic diffusion from discrete observations. *Scan. J. Statist.*, 24, (2), 211-229.

[27] Kittaneh, F. (1983) On Lipschitz functions of normal operators. *Proc. Am. Math. Soc.*, 94, 416-418.

[28] Korostelev, A. P. and Tsybakov, A.B. (1993) *Minimax theory of image reconstruction*, Lecture Notes in Statistics, vol. 82, Springer, Berlin.

[29] Kutoyants, Y.A. (1997) *On nonparametric estimation of trend coefficients in a diffusion process.* In: *Statistics and Control of Stochastic Processes*, Moscow, 230-250.

[30] Lions, P.L. and Sznitman, A.S. (1984) Stochastic differential equations with reflecting boundary conditions. *Commun. Pure Appl. Math.*, 37:, 511-537.

[31] Liptser, R. N. and Shiryaev, A. N. (2001) *Statistics of random processes. 1.: General theory. 2nd ed.* Springer, Berlin.

[32] Nussbaum, M. (1999) Minimax risk, Pinsker bound. In: *Encyclopedia of Statistical Sciences*, Update Vol. 3, 451-460. Wiley, New York.

[33] Pham, D.T. (1981) Nonparametric estimation of the drift coefficient in the diffusion equation. *Math. Operationsforsch. Statist.*, 12, (1), 61-73.

[34] Revuz, D. and Yor, M. (1999) *Continuous martingales and Brownian motion. 3rd ed.* Springer, Berlin.

[35] Rosenblatt, M. (1971) *Markov Processes. Structure and Asymptotic Behaviour.* Grundlehren der math. Wiss. 184, Springer, Berlin.

[36] Stroock, D. W. and Varadhan, S.R. (1979) *Multidimensional diffusion processes.* Grundlehren der math. Wiss. 233, Springer, Berlin.

[37] Tribouley, K. and Viennet, G. (1998) L_p adaptive density estimation in a β-mixing framework. *Ann. Inst. H. Poincaré Probab. Statist.*, 34 (1998), (2), 179-208.

[38] Viennet, G. (1997) Inequalities for absolutely regular sequences: application to density estimation. *Probab. Theory Related Fields*, 107, (4), 467–492.

[39] Yoshida, N. (1992) Estimation for diffusion processes from discrete observations. *J. Multivariate Anal.*, 41, 220-242.