stochastic processes and their applications

Stochastic Processes and their Applications 79 (1999) 135–163

Adaptive estimation in diffusion processes

Marc Hoffmann *

UPRES-A 7055, Université Paris VII, couloir 56-66 5e étage, 2 place Jussieu, 75251 Paris Cedex 05, France

Received 24 January 1997; received in revised form 4 September 1998; accepted 4 September 1998

Abstract

We study the nonparametric estimation of the coefficients of a 1-dimensional diffusion process from discrete observations. Different asymptotic frameworks are considered. Minimax rates of convergence are studied over a wide range of Besov smoothness classes. We construct estimators based on wavelet thresholding which are adaptive (with respect to an unknown degree of smoothness). The results are comparable with simpler models such as density estimation or nonparametric regression. © 1999 Published by Elsevier Science B.V. All rights reserved.

Keywords: Minimax estimation; Adaptive estimation; Diffusion processes; Discrete observations; Nonparametric regression; Wavelet orthonormal bases; Besov spaces

1. Introduction and main results

1.1. Objectives

In this paper, we investigate nonparametric estimation in 1-dimensional diffusion processes, when only a discrete sample is available. Our major point is to estimate the diffusion coefficient when one observes a diffusion process at times i/n for $i=0,\dots,n$ (Sections 1 and 2). The method we use allows to derive similar results for other diffusion models (for instance, ergodic diffusions; see Section 4).

Our goal is to carry over to diffusion models the minimax results which were obtained by Donoho et al. (1995, 1996) for the models of density estimation (with i.i.d. variables) and nonparametric regression. By comparing discretely observed diffusion models with nonparametric regression, observed on random design, we construct an estimator of the diffusion coefficient by wavelets methods. For the integrated error L_γ over compact intervals of $\mathbb{R}$ with $\gamma\in[1,\infty)$, we obtain a rate of convergence (as $n\to\infty$) which is nearly optimal in the minimax sense over a wide variety of Besov smoothness classes.

By thresholding the estimated wavelet coefficients, our procedure becomes *adaptive*: the estimator achieves the optimal rate of convergence (up to a logarithmic factor

* Tel.: +33 44 27 7974; fax: +33 4427 7674; e-mail: hoffmann@gauss.math.jussieu.fr.

0304-4149/99/$ – see front matter
PII: S0304-4149(98)00074-X

in some cases) without the knowledge of the smoothness properties of the estimated function. Such considerations are more realistic if practical purposes are considered. The results obtained show that the optimal minimax rates of convergence for the problem of estimating the diffusion coefficient are the same as in density estimation.

The nonparametric estimation of the diffusion coefficient has some history: Genon-Catalot et al. (1992) studied the estimation of the diffusion coefficient (depending on the time) in L_2 error. Florens-Zmirou (1993) studied the pointwise estimation of the diffusion coefficient (depending on the space variable) and obtained the rate $n^{-1/3}$ which is suboptimal. More recently, we obtained in Hoffmann (1996) the rate $n^{-s/(1+2s)}$ when the diffusion coefficient belongs to some ball of the Besov space $B_{sp\infty}$ (i.e. when the diffusion coefficient has "smoothness" measured in L_p of order s, see below) for the integrated error in the particular case $\gamma = p$. Similar results were obtained by Jacod (1997) using Nadaraya-Watson type estimators (both for pointwise and integrated error). This paper is thus an extension of the results in Genon-Catalot et al. (1992) and Hoffmann (1996) and also confirms that the results of Jacod (1997) (for the integrated error) are optimal.

1.2. Assumptions and construction of an estimator

(1) We consider a 1-dimensional diffusion process X defined by

$$X_t = x_0 + \int_0^t b(s, X_s)\,\mathrm{d}s + \int_0^t \sigma(X_s)\,\mathrm{d}W_s \tag{1}$$

where $x_0 \in \mathbb{R}$, W is a standard Brownian motion, b and σ are unknown. For $i = 0, \dots, 3$, let $m_i > 0$ be given constants. The assumptions on b and σ are the following.

Assumption A. The function b is continuous and $b^2(t,x) \leqslant m_0(1 + x^2)$.

Assumption B. The function σ^2 belongs to the set $\mathscr{U}$ defined by

$$\mathscr{U} = \{g \in \mathscr{C}^1(\mathbb{R}):\ m_1 \leqslant g(x) \leqslant m_2,\ \|g'\|_\infty \leqslant m_3\}$$

where we denote $\|g\|_\infty = \sup_x |g(x)|$. Assumptions A and B imply the existence of a unique (strong) solution for Eq. (1). We wish to estimate the diffusion coefficient $\sigma^2(x)$ from the observation $X^{(n)} = (X_{i/n},\ i = 0, \dots, n)$. The drift b is regarded as a nuisance parameter. In the following, we will denote by P_σ the law of X and by $\mathscr{F}_t$ the σ-field generated by $(X_s,\ 0 \leqslant s \leqslant t)$.

To estimate $\sigma^2(x)$ we require that the process X has spent some minimal time around the level x. Let

$$L_1^x = \lim_{\varepsilon \to 0} \frac{1}{2\varepsilon} \int_0^1 1_{|X_s - x| \leqslant \varepsilon}\,\mathrm{d}s$$

denote the local time of X at x up to time 1 at the Lebesgue measure scale. We introduce a level $v > 0$ (the "minimal accuracy of estimation") and define for a compact interval D

$$\mathscr{A}(v, D) = \left\{ \inf_{x \in D_\delta} L_1^x \geqslant (1+\delta)v \right\}$$

where $\delta>0$ and $D_\delta=\{x+y, x\in D, |y|\leqslant\delta\}$. The constant δ is fixed throughout the paper and is needed for technical reason. The behaviour of our estimator of $\sigma^2(x)$ will be studied conditionally on the event $\mathscr{A}(v,D)$.

(2) Let us describe our estimator $T_n(x)$ of $\sigma^2(x)$. For the reader unfamiliar with multiscale decompositions and wavelets, we recall in the appendix some basic definitions and heuristics about thresholding techniques.

Using Ito's formula, we get the following decomposition, for $i=0,\dots,n-1$

$$n(X_{(i+1)/n}-X_{i/n})^2=n\int_{i/n}^{(i+1)/n}\sigma^2(X_s)\,\mathrm{d}s+\varepsilon_{i/n}+\text{ a term of higher order} \tag{2}$$

where

$$\varepsilon_{i/n}=2n\int_{i/n}^{(i+1)/n}(X_s-X_{i/n})\sigma(X_s)\,\mathrm{d}W_s.$$

The variable $\varepsilon_{i/n}$ can be considered as a noise term when estimating $n\int_{i/n}^{(i+1)/n}\sigma^2(X_s)\,\mathrm{d}s \simeq \sigma^2(X_{i/n})$ from the observation $n(X_{(i+1)/n}-X_{i/n})^2$. The remainder term in Eq. (2) comes from a drift effect which will prove to be negligible. We thus translate our problem in a nonparametric regression setting: estimate $\sigma^2(x)$ from noisy data $n(X_{(i+1)/n}-X_{i/n})^2$ observed on the random design $(X_{i/n},\ i=0,\dots,n)$.

We need some notation. Let us be given a pair (φ,ψ) of compactly supported scaling function and wavelet such that the wavelet ψ has N vanishing moments, i.e.

$$\int x^l\psi(x)\,\mathrm{d}x=0,\quad l=0,\dots,N. \tag{3}$$

We can choose the pair in a given library of wavelets (e.g. Daubechies (1992) wavelets) and thus (φ,ψ) will be specified hereafter by N only.

The estimator T_n will depend on (φ,ψ), on the minimal time $v>0$ and on four parameters in $(0,\infty)$ which need to be tuned with the asymptotic: $(j_n, J_n, h_n, \lambda_n)$ and which are defined as follows.

For $l\in\mathbb{Z}$ we consider a grid of equispaced points x_l^n of mesh h_n, i.e. $x_l^n=lh_n$. We then define the process

$$N_i^l=\left(\sum_{0\leqslant j\leqslant i}1_{\{x_l^n\leqslant X_{j/n}<x_{l+1}^n\}}\right)\wedge\lfloor nh_nv\rfloor$$

where $\lfloor x\rfloor$ denotes the integer part of x, and an increasing sequence of random times

$$\tau_1=0 \text{ and for } i\geqslant 2\text{: } \tau_i=\inf\left\{j/n>\tau_{i-1}\text{: } \sum_{l\in\mathbb{Z}}(N_j^l-N_{\tau_{i-1}}^l)\geqslant 1\right\}\wedge 1.$$

We also need a sequence of random points ξ_{τ_i} on a finer equispaced grid of mesh $\lfloor nv\rfloor^{-1}$ defined as follows. First, set

$$l_{\tau_i}=\inf\{j\in\mathbb{Z}\text{: } X_{\tau_i}\geqslant x_j^n\},\qquad m_{\tau_i}=\#\{j\leqslant i\text{: } x_{\tau_i}^n\leqslant X_{\tau_j}<x_{\tau_i+1}^n\}$$

and let

$$\xi_{\tau_i}=x_{l_{\tau_i}-1}^n+m_{\tau_i}/\lfloor nv\rfloor.$$

Note that τ_i, l_{τ_i}, m_{τ_i} and ξ_{τ_i} depend on n. For integers (k,j) and a function g, we denote as usual $g_{jk}(x)=2^{j/2}g(2^jx-k)$. Thus, for a bounded and compactly supported g, the function g_{jk} is essentially located around $k/2^j$ in a neighbourhood of size 2^{-j}. We estimate the wavelet coefficients $c_{jk}=\int\sigma^2(x)\varphi_{jk}(x)\,\mathrm{d}x$ and $d_{jk}=\int\sigma^2(x)\psi_{jk}(x)$ of the function σ^2 by

$$\hat{c}_{jk}=\frac{1}{\lfloor nv\rfloor}\sum_{i=1}^{\lfloor nv\rfloor}\varphi_{jk}(\xi_{\tau_i})n(X_{\tau_i+1/n}-X_{\tau_i})^2, \tag{4}$$

$$\hat{d}_{jk}=\frac{1}{\lfloor nv\rfloor}\sum_{i=1}^{\lfloor nv\rfloor}\psi_{jk}(\xi_{\tau_i})n(X_{\tau_i+1/n}-X_{\tau_i})^2. \tag{5}$$

Note that $\hat{c}_{jk}$ and $\hat{d}_{jk}$ only depend on (φ,ψ), v, h_n and the data $X^{(n)}$. Finally the estimator $T_n(\cdot)=T_n(v,h_n,j_n,J_n,\lambda_n,N)(\cdot)$ at point x will be

$$T_n(x)=\sum_{k\in\mathbb{Z}}\hat{c}_{j_nk}\;\varphi_{j_nk}(x)+\sum_{j_n}^{J_n}\sum_{k\in\mathbb{Z}}\hat{d}_{jk}\;1_{\{|\hat{d}_{jk}|\geqslant\lambda_n\}}\;\psi_{jk}(x). \tag{6}$$

1.3. Main results

We will describe the quality of T_n in the *minimax theory*. For $\gamma\in[1,\infty)$ and D a compact subset of $\mathbb{R}$, let us consider the following integrated error:

$$\mathscr{L}(\gamma,T_n,D)=\int_D|T_n(x)-\sigma^2(x)|^\gamma\,\mathrm{d}x.$$

For fixed (D,γ,v) and given a set of constraint Σ, we define the *minimax risk* of T_n over Σ as

$$\mathscr{R}_{D,\gamma,v}(T_n,\alpha_n,\Sigma)=\sup_{\sigma^2\in\Sigma}E_\sigma\{\alpha_n^{-\gamma}\mathscr{L}(\gamma,T_n,D)\,|\,\mathscr{A}(v,D)\}^{1/\gamma} \tag{7}$$

where $\alpha_n>0$ is some normalizing factor. Of course, the finiteness of $\mathscr{R}_{D,\gamma,v}$ will only be meaningful if $\alpha_n\to0$ as $n\to\infty$ and the class $\Sigma\cap\mathscr{U}$ is rich enough. Our main result is about the performance of T_n when Σ is a Besov ball intersected with $\mathscr{U}$. This choice of functional constraint is standard in nonparametric estimation (see e.g. Kerkyacharian and Picard, 1993).

For $s,p,q\in[1,\infty]$, let $B_{spq}(D)$ denote the Besov space on D, i.e. the restriction of the functions of the space $B_{spq}(\mathbb{R})$ to D. Denote by $\|\cdot\|_{spq}$ the Besov norm over D (see, e.g. Cohen (1998), Meyer (1990) and/or the appendix below). For $L>0$, set

$$B_{spq}=B_{spq}(D,L)=\{g\in B_{spq}(D):\;\|g\|_{spq}\leqslant L\}.$$

Let us introduce the following quantities, involving γ, s, p and q:

$$\alpha=\frac{s}{1+2s}\wedge\frac{s-1/p+1/\gamma}{1+2s-2/p}\quad\text{and}\quad\varepsilon=sp-\frac{\gamma-p}{2}.$$

The rates of convergence are determined by α and ε: for a function in B_{spq} estimated in L_γ norm, the optimal rates are of order $(\log n/n)^\alpha$ if $\varepsilon<0$ and $n^{-\alpha}$ if $\varepsilon>0$ (see Donoho et al. (1996) and Theorems 1 and 2 below). Let

$$2^{j_n}=(n(\log n)^{\gamma-p})^{1-2\alpha}, \qquad 2^{J_n}=\left(\frac{n}{\log n}\right)^{1/(1+2s-2/p)}, \tag{8}$$

$$\lambda_n=\theta\sqrt{\frac{\log n}{n}}, \qquad h_n=n^{-s_0/(1+2s_0)} \tag{9}$$

with $s_0=s+1$ and $\theta=(1+2s-2/p)^{-1/2}K(r_0\vee(4s+2))$, where $K(\cdot)$ and r_0 are specified in Lemma 2 below. Finally, let

$$v_n=v_n(s,p,q,\gamma)=\begin{cases}\left(\dfrac{(\log n)^{1-\varepsilon/sp}}{n}\right)^\alpha & \varepsilon>0,\\[2mm] \left(\dfrac{\log n}{n}\right)^\alpha(\log n)^{(1/2-p/q\gamma)_+} & \varepsilon=0,\\[2mm] \left(\dfrac{\log n}{n}\right)^\alpha & \varepsilon<0.\end{cases}$$

Theorem 1. *Grant Assumptions* A *and* B, *and let* $v>0$ *and let* D *be a compact interval of* $\mathbb{R}$. *Let* $s>1+1/p$, $1\leqslant p\leqslant\gamma<\infty$, $q\in[1,\infty]$. *Let* T_n *be specified by* $(v,j_n,J_n,\lambda_n,h_n)$ *defined in Eqs.* (8) *and* (9) *and let* $N\geqslant\lfloor s\rfloor+1$. *We have*

$$\limsup_{n\to\infty}\mathscr{R}_{D,\gamma,v}\{T_n,v_n,B_{spq}\cap\mathscr{U}\}<\infty.$$

Our next result shows that T_n is indeed optimal (up to a logarithmic factor in some cases) for $\mathscr{R}_{D,\gamma,v}$ over $B_{spq}\cap\mathscr{U}$. Let

$$w_n=w_n(s,p,q,\gamma)=\begin{cases}\left(\dfrac{\log n}{n}\right)^\alpha & \varepsilon\leqslant 0,\\[2mm] n^{-\alpha} & \varepsilon>0.\end{cases}$$

Theorem 2. *Grant Assumptions* A *and* B, *and let* $v>0$ *and let* D *be a compact interval of* $\mathbb{R}$. *If* $s>2+1/p$, $1\leqslant p\leqslant\gamma<\infty$, $q\in[1,\infty]$, *then*

$$\liminf_{n\to\infty}\inf_{\tilde T_n}\mathscr{R}_{D,\gamma,v}\{\tilde T_n,w_n,B_{spq}\cap\mathscr{U}\}>0$$

where the infimum is taken over all estimators.

Remark. (1) We obtain the same rates as in density estimation or nonparametric regression. In the case $\varepsilon<0$, the rates v_n and w_n agree and our result is sharp. In the case $\varepsilon\geqslant 0$, we loose a logarithmic factor. The restriction $\gamma\geqslant p$ is inessential. For $\gamma<p$, usual estimators (such as those constructed in Hoffmann (1996) and Jacod (1997)) achieve the optimal rate of convergence and there is no need to introduce a thresholding procedure.

(2) The major drawback of our construction is that T_n discards some data (after the time $\tau_{\lfloor vn\rfloor}$). This has no effect on the rates of convergence, but can lead to dramatic

results for practical implementation, especially if v is too small. For technical reasons, we are unable to construct an estimator using all the data.

Let us mention that even if we knew how to do so, the criterion $\mathscr{R}_{D,v,\gamma}$ is too rough to compare procedures achieving the rate v_n: one should look at their respective *minimax efficiency*, and compare it to the optimal minimax constants for the risk $\mathscr{R}_{D,\gamma,v}$. Such results are beyond the techniques used in this paper.

(3) For $\gamma = p$ we obtain the classical rate $n^{-s/(1+2s)}$. The restriction $s > 1 + 1/p$ ensures that σ^2 is at least $\mathscr{C}^1$ over D which makes the two constraints B_{spq} and $\mathscr{U}$ compatible. The restriction $s > 2 + 1/p$ in Theorem 2 is made for technical reason.

1.4. Adaptive estimation

Fix an integer $\kappa_0 \geqslant 2$ and define

$$\mathscr{S} = \{(s,p,q):\ 1 + 1/p < s \leqslant \kappa_0,\ q \in [1,\infty],\ p \in [1,\infty)\}.$$

The goal of *adaptive estimation* in the framework of Besov smoothness classes is to construct a single estimator T_n^* (independent of s, p or q) which achieves the rate $v_n(s,p,q,\gamma)$ if σ^2 belongs to $B_{spq} \cap \mathscr{U}$ for some (s,p,q) in $\mathscr{S}$. This corresponds to more realistic considerations since the smoothness of σ^2 is generally unknown in practice.

The following results shows that a slight modification of j_n, J_n, λ_n and h_n makes T_n adaptive (up to a logarithmic term in some cases). Let

$$2^{j_n^*} = n^{1/(1+2\kappa_0)}, \qquad 2^{J_n^*} = \frac{n}{\log n}, \tag{10}$$

$$\lambda_n^* = \theta^* \sqrt{\frac{\log n}{n}}, \qquad h_n^* = n^{-\kappa_0/(1+2\kappa_0)} \tag{11}$$

with $\theta^* = (1+2\kappa_0)^{-1/2}\tilde{K}(2\kappa_0+4)$. The function $\tilde{K}(\cdot)$ is obtained from $K(\cdot)$ by specifying $\Gamma = 1/4$ in Lemma 2 (see below). Thus θ^* only depends on v, ψ, γ and m_2. Define

$$z_n = z_n(s,p,q,\gamma) = \begin{cases} \left(\dfrac{\log n}{n}\right)^{\alpha} & \varepsilon \neq 0 \\ \left(\dfrac{\log n}{n}\right)^{\alpha} (\log n)^{(1/2 - p/q\gamma)_+} & \varepsilon = 0 \end{cases}$$

and note that z_n differs from the optimal rate only by a logarithmic factor in some cases (compare with Theorem 2 above).

Theorem 3. *Grant Assumptions* A *and* B, *and let* $v > 0$ *and* D *a compact interval of* $\mathbb{R}$. *Let* T_n^* *be specified by* $(v, j_n^*, J_n^*, \lambda_n^*, h_n^*)$ *defined by Eqs.* (10) *and* (11) *and let* $N \geqslant r_0 + 1$.

Then T_n^* *is* (*nearly*) *adaptive over* $\mathscr{S}$, *i.e.*

$$\sup_{(s,p,q)\in\mathscr{S}} \limsup_{n\to\infty} \mathscr{R}_{D,\gamma,v}\{T_n^*, z_n, B_{spq} \cap \mathscr{U}\} < \infty.$$

Note that T_n^* does not depend on (s,p,q).

2. Proofs

2.1. Proof of Theorem 1

The proof of Theorem 1 essentially relies on two kinds of inequalities: a moment bound for the difference $\hat{c}_{jk} - c_{jk}$ and an exponential inequality for the deviation of $\hat{d}_{jk} - d_{jk}$. Given these tools, the proof of Theorem 1 is obtained in the same line as Theorem 3 in Donoho et al. (1996). Nevertheless, for the reader's convenience, we will give a self-containing proof of Theorem 1.

2.1.1. Preliminary lemmas

Let $\tilde{P}_\sigma$ denote the law of the process Y such that $\mathrm{d}Y_t = \sigma(Y_t)\,\mathrm{d}W_t$, $Y_0 = x_0$. For a compactly supported function g we denote by $\mathscr{S}(g)$ the support of g. In this section, whenever reference is made to j, we assume that $j_n \leqslant j \leqslant J_n$. We denote by C a generic constant, possibly varying from line to line, which may depend on (φ, ψ) and m_i, $i = 0, \dots, 4$. Any other dependence will be explicitly mentioned.

Lemma 1. *For $r \in [1, \infty)$ we have*

$$\tilde{E}_\sigma\{|\hat{c}_{jk} - c_{jk}|^r \mid \mathscr{A}(v, D)\} \leqslant C(v, D, r) n^{-r/2}. \tag{12}$$

Proof. The proof will be done in several steps. We write

$$\hat{c}_{jk} - c_{jk} = A_{jk}^{(n)} + B_{jk}^{(n)} + C_{jk}^{(n)}$$

where

$$A_{jk}^{(n)} = \frac{1}{\lfloor nv \rfloor} \sum_{i=1}^{\lfloor nv \rfloor} \sigma^2(X_{\tau_i}) \varphi_{jk}(\xi_{\tau_i}) - c_{jk},$$

$$B_{jk}^{(n)} = \frac{1}{\lfloor nv \rfloor} \sum_{i=1}^{\lfloor nv \rfloor} \int_{\tau_i}^{\tau_i + 1/n} (\sigma^2(X_s) - \sigma^2(X_{\tau_i})) \varphi_{jk}(\xi_{\tau_i})\,\mathrm{d}s,$$

$$C_{jk}^{(n)} = \frac{1}{\lfloor nv \rfloor} \sum_{i=1}^{\lfloor nv \rfloor} \varphi_{jk}(\xi_{\tau_i}) \varepsilon_{\tau_i}.$$

Recall from Section 1 that $\varepsilon_{\tau_i} = 2n \int_{\tau_i}^{\tau_i + 1/n} (X_s - X_{\tau_i}) \sigma(X_s)\,\mathrm{d}W_s$ and that we adopt the convention $\varepsilon_1 = 0$. We will successively give moment bounds for $A_{jk}^{(n)}$, $B_{jk}^{(n)}$ and $C_{jk}^{(n)}$.

(a) Let ϕ be a smooth function with support in $[-\frac{1}{2}, \frac{1}{2}]$ such that $\|\phi\|_\infty \leqslant 1$ and $\int \phi(x)\,\mathrm{d}x \geqslant \frac{1}{2}$. Set $\bar{\phi}(x) = (\int \phi(u)\,\mathrm{d}u)^{-1} \phi(x)$. Let $v_0 = v(\int \phi(u)\,\mathrm{d}u)^{-1}$ and

$$L_n(x) = \frac{1}{n h_n} \sum_{i=0}^{n-1} \bar{\phi}(h_n^{-1}(X_{i/n} - x)),$$

$$\mathscr{T}_{jk} = \bigcap_{l:\ x_n^l \in \mathscr{S}(\varphi_{jk})} \{L_n(x_n^l) \geqslant v_0\}.$$

Note that $L_n(x)$ is an approximation kernel of the local time L_1^x. We have the following inclusion

$$(L_n(x_l^n)\geqslant v_0)=\left(\frac{1}{nh_n}\sum_{i=0}^{n-1}\phi\left(h_n^{-1}(X_{i/n}-x_l^n)\right)\geqslant v\right)\subseteq\left(\sum_{i=0}^{n-1}1_{|X_{i/n}-x_l^n|\leqslant h_n}\geqslant\lfloor nh_n v\rfloor\right).$$

It follows that $\tau_i<1$ on $\mathscr{T}_{jk}$, hence, from the construction of τ_i and ξ_{τ_i}:

$$|\sigma^2(X_{\tau_i})-\sigma^2(\xi_{\tau_i})|1_{\mathscr{T}_{jk}}\leqslant C|X_{\tau_i}-\xi_{\tau_i}|1_{\tau_i<1}\leqslant Ch_n. \tag{13}$$

(b) Write now $A_{jk}^{(n)}-c_{jk}=R_{k,1}^{(n)}+R_{k,2}^{(n)}$, where

$$R_{k,1}^{(n)}=\frac{1}{\lfloor nv\rfloor}\sum_{i=1}^{\lfloor nv\rfloor}(\sigma^2(X_{\tau_i})-\sigma^2(\xi_{\tau_i}))\varphi_{jk}(\xi_{\tau_i}),$$

$$R_{k,2}^{(n)}=\frac{1}{\lfloor nv\rfloor}\sum_{i=1}^{\lfloor nv\rfloor}\sigma^2(\xi_{\tau_i})\varphi_{jk}(\xi_{\tau_i})-c_{jk}.$$

Using that on $\mathscr{T}_{jk}$, the sum in $A_{jk}^{(n)}$ has of order $n2^{-j}$ terms together with Eq. (13) yields

$$\tilde{E}_\sigma^2(|R_{k,1}^{(n)}|^r|\mathscr{T}_{jk})\leqslant C(v,r)2^{-jr/2}h_n^r.$$

Using Riemann's approximation, we also have $\tilde{E}_\sigma(|R_{k,2}^{(n)}|^r|\mathscr{T}_{jk})\leqslant C(v,r)n^{-r}2^{jr/2}$. From the choice of j_n and J_n, we derive $E_\sigma(|A_{jk}^{(n)}-c_{jk}|^r|\mathscr{T}_{jk})\leqslant C(v,r)n^{-r/2}$.

(c) Note that $|A_{jk}^{(n)}-c_{jk}|\leqslant C2^{j/2}$, hence

$$\begin{aligned}\tilde{E}_\sigma\{|A_{jk}^{(n)}-c_{jk}|^r|\mathscr{A}(v,D)\} &\leqslant C(v,D)(\tilde{E}_\sigma\{|Q_1|^r\cap\mathscr{T}_{jk}\}) \\ &\quad +(2^{j/2}\tilde{P}_\sigma\{\mathscr{T}_{jk}^c\cap\mathscr{A}(v,D)\}).\end{aligned} \tag{14}$$

It remains to study the second term in the right-hand side of Eq. (14). Clearly

$$\begin{aligned}\tilde{P}_\sigma\left(\mathscr{T}_{jk}^c\bigcap\mathscr{A}(v,D)\right) &\leqslant \sum_{l:x_n^l\in\mathscr{S}(\varphi_{jk})}\tilde{P}_\sigma(\{L_1^{x_n^l}\geqslant(1+\delta)v\}\cap\{L_n(x_n^l)<v\}) \\ &\leqslant Ch_n^{-1}2^{-j}\sup_{x\in\mathscr{S}(\varphi_{jk})}\tilde{P}_\sigma(|L_1^x-L_n(x)|\geqslant\delta v).\end{aligned}$$

By approximation of the local time, this last term is of arbitrarily small order (in power of n), therefore $\tilde{E}_\sigma\{|A_{jk}^{(n)}-c_{jk}|^r|\mathscr{A}(v,D)\}\leqslant C(v,D)n^{-r/2}$.

(d) Using the Burckholder–Davis–Gundy inequality

$$\tilde{E}_\sigma\left(\left|\int_{\tau_i}^{\tau_i+1/n}[\sigma^2(X_s)-\sigma^2(X_{\tau_i})]\mathrm{d}s\right|^r\right)\leqslant C(r)n^{-r/2}.$$

The same penalty argument by the event $\mathscr{T}_{jk}$ as in (b) and (c) ensures that the sum in $B_{jk}^{(n)}$ has at most of order $n2^{-j}$ terms. We derive $\tilde{E}_\sigma(|B_{jk}^{(n)}|^r\mid\mathscr{A}(v,D))\leqslant C(v,D,r)$ $2^{-j/2}n^{-r/2}$.

(e) For the last term, we use a martingale version of Rosenthal inequality (see Hall and Hegde, 1980, p. 23). We have

$$\tilde{E}_\sigma(|B^n_{jk}|^r) \leqslant C(r,v)\left\{\left(\tilde{E}_\sigma\left\{\sum_{i=0}^{\lfloor nv\rfloor}\tilde{\varphi}^2_{jk}(\xi_{\tau_i})\tilde{E}_\sigma(\varepsilon^2_{n\tau_i,n}|\mathscr{F}_{\tau_i})\right\}\right)^{r/2} + \sum_{i=0}^{\lfloor nv\rfloor}\tilde{E}_\sigma(|\varphi_{jk}(\xi_{\tau_i})\varepsilon_{n\tau_i,n}|^r)\right\}.$$

From the Burckholder–Davis–Gundy inequality, it is easily seen that

$$\forall r\in[1,\infty):\ \tilde{E}_\sigma(|\varepsilon_{n\tau_i,n}|^r \mid \mathscr{F}_{\tau_i})\leqslant C(r)2^{jr/2}.$$

Hence, using that the ξ_{τ_i} are all distinct on $(\tau_{\lfloor nh_n v\rfloor}<1)$ and that $\varepsilon_{\tau_i}=0$ on $(\tau_i\geqslant 1)$, we see that the sum in $C^{(n)}_{jk}$ has at most of order $n2^{-j}$ terms. We easily derive $\tilde{E}_\sigma(|C^{(n)}_{jk}|^r)\leqslant C(r,v)(n^{-r/2}+n^{-r+1}2^{jr/2})$. From the choice of j_n and J_n, we obtain the desired bound, hence (15) follows and Lemma 1 is proved.

Lemma 2. *Let*

$$\Gamma=\frac{1}{2}\left(1-\frac{1+2s}{1+2s_0}\right).$$

Let $r_0=4|\mathscr{S}(\psi)|(3c_0\Gamma\log 2)^{-1}$. *For* $r\geqslant r_0$, *the choice* $K=12r\Gamma v^{-1}c_*^2m_2\|\psi\|_\infty$ *yields the following inequality*:

$$\tilde{P}_\sigma\{|\hat{d}_{jk}-d_{jk}|\geqslant K(r)\sqrt{j/n}|\mathscr{A}(v,D)\}\leqslant C(v,D,r)2^{-j\Gamma r}. \tag{15}$$

Proof. We use the following decomposition.

$$\hat{d}_{jk}-d_{jk}=U^{(n)}_{jk}+V^{(n)}_{jk}$$

where

$$U^{(n)}_{jk}=\frac{1}{\lfloor nv\rfloor}\sum_{i=1}^{\lfloor nv\rfloor}\left(n\int_{\tau_i}^{\tau_i+1/n}\sigma^2(X_s)\mathrm{d}s\right)\psi_{jk}(\xi_{\tau_i})-d_{jk},$$

$$V^{(n)}_{jk}=\frac{1}{\lfloor nv\rfloor}\sum_{i=1}^{\lfloor nv\rfloor}\varepsilon_{\tau_i}\psi_{jk}(\xi_{\tau_i}).$$

(a) We first give a bound for

$$\tilde{P}_\sigma\left\{|U^{(n)}_{jk}|\geqslant\frac{1}{2}K\sqrt{j/n}\,\middle|\,\mathscr{A}(v,D)\right\}$$

where $K>0$ is to be determined later. Replacing φ by ψ in (b)–(d) of Lemma 1, we obtain that for $r\geqslant 1$

$$\tilde{E}_\sigma\{|U^{(n)}_{jk}|^r\}\leqslant C(r,v,D)\{2^{-jr/2}h_n^r+2^{jr/2}(n^{-r}+n^{-r/2})\}$$

which is less than $C(r,v,D)2^{-jr/2}h_n^r$ for $j_n \leqslant j \leqslant J_n$. Chebyshev's inequality yields

$$\tilde{P}_\sigma\left\{|U_{jk}^{(n)}| \geqslant \frac{1}{2}K\sqrt{j/n}\,\middle|\, \mathscr{A}(v,D)\right\} \leqslant C(r,v,D)\left(\frac{2}{K}\right)^r n^{r/2}h_n^r j^{-r/2}2^{-jr/2}$$
$$\leqslant C(r,v,D)n^{r/2(1+2s_0)}2^{-jr/2} \tag{16}$$

where we used that $h_n = n^{-s_0/1+2s_0}$. Writing $2^{-jr/2} = 2^{-j\Gamma r}2^{-j(1/2-\Gamma)r}$, using that for $j_n \leqslant j \leqslant J_n$, we have

$$2^{-j} \leqslant n^{2\alpha-1} \leqslant n^{2s/(1+2s)-1} = n^{-(1/1+2s)}$$

and inserting $(2/K)^r$ in the constant $C(r,v,D)$ of Eq. (16), we derive that the right-hand side of Eq. (16) is less than

$$C(r,v,D,K)2^{-jr/2}2^{-j\Gamma r}n^{r(1/2(1+2s_0)-1/(1+2s)(\frac{1}{2}-\Gamma))}.$$

From the choice of Γ, the term in power of n vanishes and we obtain

$$\tilde{P}_\sigma\left\{|U_{jk}^{(n)}| \geqslant \frac{1}{2}K\sqrt{j/n}\,\middle|\, \mathscr{A}(v,D,K)\right\} \leqslant C(v,r,D)2^{-j\Gamma r}. \tag{17}$$

(b) We now turn to $V_{jk}^{(n)}$. The following Bernstein type inequality will be needed

Lemma 3. *Let $(M_n, n \geqslant 0)$ be a $(\mathscr{F}_n)$-martingale such that for all $k \in [2,\infty), E(|M_{i+1} - M_i|^k|\mathscr{F}_i) \leqslant c^k k^k$ for some constant $c > 0$. Then*

$$\forall t \geqslant 0: \quad P(|M_n| \geqslant t) \leqslant 2\exp\left(-\frac{t^2}{ce(2cn+t)}\right).$$

Proof. Pick up $u \in (0, 1/ce)$. We have

$$\sum_{k=2}^{\infty}\frac{u^k}{k!}E(|M_{i+1} - M_i|^k|\mathscr{F}_i) \leqslant (uc)^2\sum_{k=2}^{\infty}\frac{u^k}{k!}(uc)^{k-2}. \tag{18}$$

Since $k! \geqslant k^k e^{-k+1}$ for $k \geqslant 2$ (apply the concavity of the logarithm) the last term in Eq. (18) is less than $(uc)^2 e\sum_{k=2}^{\infty}(uce)^{k-2} = (uc)^2 e/(1-uce)$. Using $1 + x \leqslant \exp x$, it follows that

$$E(e^{\lambda(M_{i+1}-M_i)}|\mathscr{F}_i) \leqslant 1 + \sum_{k=2}^{\infty}\frac{u^k}{k!}E(|M_{i+1} - M_i|^k|\mathscr{F}_i) \leqslant \exp\frac{(uc)^2 e}{(1-uce)}. \tag{19}$$

For $t \geqslant 0$, using Chebyshev's exponential inequality and Eq. (19), we get

$$P(M_n \geqslant t) \leqslant e^{-ut}E(e^{uM_n}) \leqslant e^{-ut}\exp\left(n\frac{(uc)^2 e}{(1-uce)}\right),$$

where the last inequality is obtained by conditioning n times w.r.t. $\mathscr{F}_i$, $i = 1,\ldots,n$. The choice $u = t/(2nc^2 e + cet)$ yields

$$P(M_n \geqslant t) \leqslant \exp\left(-\frac{1}{2}\frac{t^2}{ce(2cn+t)}\right). \tag{20}$$

It suffices to note that $P(|M_n|\geqslant t)\leqslant P(M_n\geqslant t)+P(M_n\leqslant -t)$ and to apply (20) to $-M_n$ to obtain Lemma 3. □

Proof of Lemma 2 (*Completion*)**.** We apply Lemma 3 with $M_{i+1}-M_i=\varepsilon_{\tau_i}\psi_{jk}(\xi_{\tau_i})$. First note that

$$|M_{i+1}-M_i|\leqslant\|\psi\|_\infty 2^{j/2}|\varepsilon_{\tau_i}|.$$

Let $k\geqslant 2$. Using twice the Burckholder–Davis–Gundy inequality with best constant $C_k=c_*\sqrt{k}$ (see for instance Barlow and Yor, 1982) together with Jensen inequality, we successively have, for $1\leqslant i\leqslant\lfloor nv\rfloor$

$$\begin{aligned}\tilde{E}_\sigma\{|\varepsilon_{\tau_i}|^k|\mathscr{F}_{\tau_i}\} &= 2^k n^k \tilde{E}_\sigma\left\{\left|\int_{\tau_i}^{\tau_i+1/n}(X_s-X_{\tau_i})\sigma(X_s)\mathrm{d}W_s\right|^k\,\middle|\,\mathscr{F}_{\tau_i}\right\}\\ &\leqslant 2^k n^p k c_*^k k^{k/2}\tilde{E}_\sigma\left\{\left(\int_{\tau_i}^{\tau_i+1/n}(X_s-X_{\tau_i})^2\sigma^2(X_s)\mathrm{d}s\right)^{k/2}\,\middle|\,\mathscr{F}_{\tau_i}\right\}\\ &\leqslant 2^k n^{k/2+1}c_*^k k^{k/2}\|\sigma^2\|_\infty^{k/2}\int_{\tau_i}^{\tau_i+1/n}\tilde{E}_\sigma\{|X_s-X_{\tau_i}|^k\mid\mathscr{F}_{\tau_i}\}\\ &\leqslant 2^k c_*^{2k}k^k\|\sigma^2\|_\infty\leqslant(2c_*^2 m_2)^k.\end{aligned}$$

In conclusion, we can apply Lemma 3 to M_i with $c=2^{j/2}c_0$, where

$$c_0=2c_*^2 m_2\|\psi\|_\infty.$$

We obtain

$$\begin{aligned}&\tilde{P}_\sigma\left\{|V_{jk}^{(n)}|\geqslant\frac{1}{2}K\sqrt{j/n}\,\middle|\,\mathscr{A}(v,D)\right\}\\ &\quad\leqslant 2C(v,D)\exp\left\{-\frac{K^2n^2v^2j}{2^{j/2+1}c_0[42^{-j/2}n|\mathscr{S}(\psi)|+k\sqrt{jnv}]}\right\}\end{aligned}$$

where we used that the sum in i has at most $\lfloor nv2^{-j}|\mathscr{S}(\psi)|\rfloor$ terms. Using that for $j_n\leqslant j\leqslant J_n$, we have $j2^j\leqslant n$, the right-hand side in the previous inequality is less than

$$C(v,D)\exp\{-j\vartheta_r\},\tag{21}$$

where

$$\vartheta_r=\frac{k^2v^2}{2c_0[4|\mathscr{S}(\psi)|+Kv]}.$$

We look for a K which provides the following inequality:

$$\Gamma r-\frac{\vartheta_r}{\log 2}\leqslant 0.\tag{22}$$

Elementary computation shows that for $r\geqslant r_0=4|\mathscr{S}(\psi)|(3c_0\Gamma\log 2)^{-1}$, the choice

$$K=K(r,v,\Gamma,\psi,m_2)=12r\Gamma v^{-1}c_*^2m_2\|\psi\|_\infty$$

yields Eq. (22). This together with Eq. (21) entails

$$\tilde{P}_\sigma \left\{ |V_{jk}^{(n)}| \geqslant \left| \frac{1}{2} K \sqrt{j/n} \right| \mathscr{A}(v,D) \right\} \leqslant C(v,r,D) 2^{-j\Gamma r}. \tag{23}$$

Putting together Eqs. (17) and (23) proves Lemma 2. □

2.1.2. *Proof of Theorem 1*

For $x \in \mathbb{R}$, we have

$$\sigma^2(x) = \sum_k c_{j_n k} \varphi_{j_n k}(x) + \sum_{j=j_n}^{J_n} \sum_k d_{jk} \psi_{jk}(x) + \left[\sigma^2(x) - \sum_k c_{J_n k} \varphi_{J_n k}(x) \right].$$

Let us stress that since φ and ψ are compactly supported, the sum in k is finite hence the above decomposition is well defined. Let us write

$$T_n(x) - \sigma^2(x) = A_n(x) + B_n(x) + C_n(x)$$

where

$$A_n(x) = \sum_k (\hat{c}_{j_n k} - c_{j_n k}) \varphi_{j_n k}(x),$$

$$B_n(x) = \sum_{j=j_n}^{J_n} \sum_k (\hat{d}_{jk} 1_{|\hat{d}_{jk}| \geqslant \lambda_n} - d_{jk}) \psi_{jk}(x),$$

$$C_n(x) = \sigma^2(x) - \sum_k c_{J_n k} \varphi_{J_n k}(x).$$

Clearly

$$\mathscr{R}_{D,\gamma,v}(T_n, v_n, B_{spq} \cap \mathscr{U}) \leqslant C(v,D,\gamma) \{ \mathscr{R}_n^{(1)} + \mathscr{R}_n^{(2)} + \mathscr{R}_n^{(3)} \} \tag{24}$$

where

$$\mathscr{R}_n^{(1)} = \sup_{\sigma^2 \in \mathscr{U}} E_\sigma \left(v_n^{-\gamma} \int_D |A_n(x)|^\gamma \mathrm{d}x \,\middle|\, \mathscr{A}(v,D) \right)^{1/\gamma},$$

$$\mathscr{R}_n^{(2)} = \sup_{\sigma^2 \in \mathscr{U}} E_\sigma \left(v_n^{-\gamma} \int_D |B_n(x)|^\gamma \mathrm{d}x \,\middle|\, \mathscr{A}(v,D) \right)^{1/\gamma},$$

$$\mathscr{R}_n^{(3)} = \sup_{\sigma^2 \in B_{spq}} \left(v_n^{-\gamma} \int_D |C_n(x)|^\gamma \, \mathrm{d}x \right)^{1/\gamma}.$$

(a) For $\mathscr{R}_n^{(1)}$, using the localization property of φ, we have (see e.g. Meyer, 1990, p. 30)

$$E_\sigma \left| \left(\int_D |A_n(x)|^\gamma \, \mathrm{d}x \,\middle|\, \mathscr{A}(v,D) \right) \right. \leqslant C(\gamma) 2^{j_n(\gamma/2-1)} \sum_k E_\sigma(|\hat{c}_{j_n k} - c_{j_n k}|^\gamma \mid \mathscr{A}(v,D)). \tag{25}$$

The measures P_σ and $\tilde{P}_\sigma$ are equivalent on $\mathscr{F}_1$ with density

$$\mathscr{D} = \frac{\mathrm{d}P_\sigma}{\mathrm{d}\tilde{P}_\sigma} = \exp\left(\int_0^1 \frac{b(s,X_s)}{\sigma^2(X_s)}\,\mathrm{d}X_s - \frac{1}{2}\int_0^1 \frac{b^2(s,X_s)}{\sigma^2(X_s)}\,\mathrm{d}s \right).$$

By Hölder inequality, it follows that

$$\begin{aligned} E_\sigma(|\hat{c}_{j_nk} - c_{j_nk}|^\gamma \mid \mathscr{A}(v,D)) &\leqslant \tilde{E}_\sigma(\mathscr{D}\,|\hat{c}_{j_nk} - c_{j_nk}|^\gamma \,|\mathscr{A}(v,D)) \\ &\leqslant \tilde{E}_\sigma(\mathscr{D}^u)^{1/u}\tilde{E}_\sigma(|\hat{c}_{j_nk} - c_{j_nk}|^{v\gamma}|\mathscr{A}(v,D))^{1/v} \end{aligned}$$

where $u^{-1} + v^{-1} = 1$. The second term in the right-hand side of the last inequality is less than $C(v,D,v,\gamma)n^{-\gamma/2}$ by Lemma 1. For the first term, it suffices to pick $u \in (1,\infty)$ such that

$$\sup_{\sigma^2\in\mathscr{U}} \tilde{E}_\sigma(\mathscr{D}^u) < \infty$$

a choice which is possible since Assumption A implies the existence of some $\tau > 0$ such that $\sup_{0\leqslant t\leqslant 1} E_\sigma[e^{\tau X_t^2}] < \infty$ (see e.g. Theorem 4.7. in Lipster and Shiryayev (1977)). This yields

$$\mathscr{R}_n^{(1)} \leqslant C(v,D,\gamma)v_n^{-1}\left(\frac{2^{j_n}}{n}\right)^{1/2}.$$

From the choice of j_n, this last term is bounded.

(b) Let us study $\mathscr{R}_n^{(3)}$. Consider a function $\bar{\sigma}^2 \in B_{spq}(\mathbb{R})$ such that

$$\bar{\sigma}^2(x) = \begin{cases} \sigma^2(x) & \text{if } x \in D_\delta, \\ 0 & \text{if } x \in \mathbb{R}\backslash D_\delta, \end{cases}$$

and $\|\bar{\sigma}^2\|_{spq} \leqslant L + \delta$, a choice which is obviously possible. Let P_j denote the projection operator onto $V_j = \mathrm{span}\{\varphi_{jk},\ k \in \mathbb{Z}\}$. We have

$$\int_D |C_n(x)|^\gamma \mathrm{d}x \leqslant \int_{D_\delta} |\bar{\sigma}^2(x) - P_{J_n}\bar{\sigma}^2(x)|^\gamma \mathrm{d}x = \|\bar{\sigma}^2 - P_{J_n}\bar{\sigma}^2\|^\gamma_{L^\gamma(\mathbb{R})}.$$

Using the Sobolev embedding $B_{spq}(\mathbb{R}) \subset B_{\bar{s}\gamma\infty}(\mathbb{R})$ where

$$\bar{s} = s - \frac{1}{p} + \frac{1}{\gamma}$$

and the rate of approximation provided by P_{J_n} (see e.g. Appendix), the last quantity is less than

$$C(\gamma)(L+\delta)^\gamma 2^{-J_n\bar{s}\gamma}.$$

From the choice of J_n, this bound has the right order.

(c) We turn to the main term $\mathscr{R}_n^{(2)}$. We will use the following decomposition:

$$B_n(x) = \sum_{i=1}^{4}\sum_{j=j_n}^{J_n}\sum_k h_{jk}^{(i)}\psi_{jk}(x)$$

where

$$h_{jk}^{(1)} = (\hat{d}_{jk} - d_{jk})1_{\{|\hat{d}_{jk}|\geqslant\lambda_n, |d_{jk}|<\lambda_n/2\}},$$
$$h_{jk}^{(2)} = (\hat{d}_{jk} - d_{jk})1_{\{|\hat{d}_{jk}|\geqslant\lambda_n, |d_{jk}|\geqslant\lambda_n/2\}},$$
$$h_{jk}^{(3)} = -d_{jk}1_{\{|\hat{d}_{jk}|\leqslant\lambda_n, |d_{jk}|>2\lambda_n\}},$$
$$h_{jk}^{(4)} = -d_{jk}1_{\{|\hat{d}_{jk}|\leqslant\lambda_n, |d_{jk}|\leqslant 2\lambda_n\}}.$$

For notational simplicity, we will abbreviate $P_\sigma(\cdot|\mathscr{A}(v,D))$ by $\bar{P}_\sigma$. For $t\geqslant 0$, set $S_n(t)=\sum_{j=j_n}^{J_n} 2^{jt}$. The following bound will be repeatedly used in this sequel

Lemma 4. *For $i=1,\ldots,4$ and $\beta\geqslant 0$*

$$\bar{E}_\sigma\left\{\int_D\left|\sum_{j=j_n}^{J_n}\sum_k h_{jk}^{(i)}\psi_{jk}(x)\right|^\gamma \mathrm{d}x\right\}$$
$$\leqslant C(\gamma)\sum_{j=j_n}^{J_n} 2^{j(\gamma/2-1)}\sum_k \bar{E}_\sigma(|h_{jk}^{(i)}|^\gamma) \quad 1\leqslant\gamma\leqslant 2$$
$$\leqslant C(\gamma)S_n\left(\beta\frac{\gamma}{\gamma-2}\right)^{\gamma/2-1}\sum_{j=j_n}^{J_n} 2^{j(\gamma/2-1-\beta\gamma/2)}\sum_k \bar{E}_\sigma(|h_{jk}^{(i)}|^\gamma) \quad \gamma>2.$$

Proof. Straightforward using the localization of ψ and Hölder inequality. □

(c1) Note that from the construction of θ, we have $K\sqrt{j/n}\leqslant\lambda_n$. Using Hölder inequality, the inclusion $\{|\hat{d}_{jk}|\geqslant\lambda_n, |d_{jk}|<\lambda_n/2\}\subset\{|\hat{d}_{jk}-d_{jk}|>\lambda_n/2\}$, and Lemmas 1 and 2, we successively have, for $r\geqslant r_0$

$$\sum_k \bar{E}_\sigma(|h_{jk}^{(1)}|^\gamma) \leqslant \sum_k \bar{E}_\sigma(|\hat{d}_{jk}-d_{jk}|^{\gamma u})^{1/u}\bar{P}_\sigma(|\hat{d}_{jk}-d_{jk}|>\lambda_n/2)^{1/v}$$
$$\leqslant C(\gamma,u,r)n^{-\gamma/2}2^{j(1-r\Gamma/v)}$$

where $u^{-1}+v^{-1}=1$. Applying Lemma 4 yields

$$\bar{E}_\sigma\left\{\int_D\left|\sum_{j=j_n}^{J_n}\sum_k h_{jk}^{(1)}\psi_{jk}(x)\right|^\gamma \mathrm{d}x\right\}$$
$$\leqslant C(v,D,\gamma,u,r)n^{-\gamma/2}S_n\left(\beta\frac{\gamma}{\gamma-2}\right)^{(\gamma/2-1)_+}\sum_{j=j_n}^{J_n} 2^{j(\gamma/2-1-\beta\gamma/2)}2^{j(1-\Gamma ur)}$$
$$\leqslant C(v,D,\gamma,u,r)n^{-\gamma/2}S_n\left(\beta\frac{\gamma}{\gamma-2}\right)^{(\gamma/2-1)_+} S_n\left((1-\beta)\gamma/2-\Gamma vr\right). \tag{26}$$

When $\gamma>2$ we have $S_n(\beta\gamma/(\gamma-2))^{(\gamma/2-1)_+}S_n(t)\leqslant C(t,\beta,\gamma)2^{(\beta\gamma/2+t)j_s}$, where $j_s=j_n$ if both $\beta>0$ and $t<0$ and $j_s=J_n$ if β and $t>0$. When $\gamma\leqslant 2$, $S(t)\leqslant C(t)2^{tj_s}$, with $j_s=j_n$ if $t<0$ and $j_s=J_n$ if $t>0$.

Since r can be chosen arbitrarily large and the choice of β is free, when $\gamma>2$, we can choose the appropriate argument of S_n in Eq. (26) to be negative (both arguments when $\gamma>2$ and the second one when $1\leqslant\gamma\leqslant 2$). Thus, for $\gamma\geqslant 1$

$$\bar{E}_\sigma\left\{\int_D\left|\sum_{j=j_n}^{J_n}\sum_k h_{jk}^{(1)}\psi_{jk}(x)\right|^\gamma \mathrm{d}x\right\}\leqslant C(v,D,\gamma,)2^{j_n(\gamma/2-vr)}n^{-\gamma/2}.$$

For any choice of $r\geqslant r_0$, this bound is asymptotically negligible w.r.t. v_n^γ.

(c2) For $h_{jk}^{(3)}$, we still have the inclusion of (c1). Therefore, for $r\geqslant r_0$

$$\begin{aligned}\sum_k \bar{E}_\sigma(|\hat{h}_{jk}^{(3)}|^\gamma) &\leqslant \sum_k |d_{jk}|^\gamma \bar{P}_\sigma(|\hat{d}_{jk}-d_{jk}|>\lambda_n/2)\\ &\leqslant \|d_{j.}\|_\gamma^\gamma 2^{-j\Gamma r}\end{aligned}\tag{27}$$

where we also denote by $\|\cdot\|_\gamma$ the l^γ norm on sequence spaces. We then use the inclusion $B_{\bar{s}\gamma q}(\mathbb{R})\subset B_{\bar{s}\gamma\infty}(\mathbb{R})$ and the norm equivalence

$$\sup_{j\leqslant 0}(2^{j(\bar{s}+1/2-1/\gamma)}\|d_{j.}\|_\gamma)\asymp\left\|\sum_k h_{.k}^{(3)}\psi_{.k}\right\|_{\bar{s}\gamma\infty}$$

to derive

$$\|d_{j.}\|_\gamma\leqslant C(\gamma,s)\left\|\sum_k h_{.k}^{(3)}\psi_{.k}\right\|_{\bar{s}\gamma\infty}^\gamma 2^{-j(\bar{s}\gamma+\gamma/2-1+r)}.\tag{28}$$

Finally choosing β and r as for $h_{jk}^{(1)}$ and using the embedding $B_{sp\infty}(\mathbb{R})\subset B_{\bar{s}\gamma\infty}(\mathbb{R})$ we derive

$$\begin{aligned}&\bar{E}_\sigma\left(\int_D\left|\sum_{j=j_n}^{J_n}\sum_k h_{jk}^{(3)}\psi_{jk}(x)\right|^\gamma \mathrm{d}x\right)\\ &\leqslant C(v,D,\gamma)S_n\left(\beta\frac{\gamma}{\gamma-2}\right)^{(\gamma/2-1)_+}S_n(-\gamma(\beta/2+\bar{s})-r)\\ &\leqslant C(v,D,\gamma)2^{-j_n(r+\bar{s}\gamma)}.\end{aligned}$$

This term also shows to be asymptotically negligible by taking r large enough.

(c3) Define

$$\mathscr{K}_j(g)=\{k\in\mathbb{Z}\colon\ \mathscr{S}(g_{jk})\subset(1+\delta)D\}$$

and let $\mathscr{B}_j=\{k\in\mathscr{K}_j(\psi)\colon\ |d_{jk}|>\lambda_n/2\}$. Using Lemma 1, we get

$$\begin{aligned}\sum_k\bar{E}_\sigma\left(|\hat{h}_{jk}^{(2)}|^\gamma\right) &\leqslant C(v,D,\gamma)n^{-\gamma/2}\sum_{k\in\mathscr{B}_j}|2d_{jk}\lambda_n|^p\\ &\leqslant C(v,D,\gamma)\|d_{j.}\|_p^p j^{-p/2}n^{-(\gamma-p)/2}\\ &\leqslant C(v,D,\gamma)\left\|\sum_k h_{jk}^{(2)}\psi_{jk}\right\|_{sp\infty}^p 2^{-j(s+1/2-1/p)p}j^{-p/2}n^{(\gamma-p)/2}.\end{aligned}\tag{29}$$

In the case $\varepsilon \neq 0$, we use Lemma 4 and we obtain

$$\bar{E}_\sigma\left(\int_D \left|\sum_{j=j_n}^{J_n}\sum_k h_{jk}^{(2)}\psi_{jk}(x)\right|^\gamma \mathrm{d}x\right)$$

$$\leqslant C(v,D,\gamma)n^{-(\gamma-p)/2}S_n\left(\beta\frac{\gamma}{\gamma-2}\right)^{(\gamma/2-1)_+}S_n(-\varepsilon-\beta\gamma/2)$$

$$\leqslant C(v,D,\gamma)n^{-(\gamma-p)/2}2^{\max(-J_n\varepsilon,-j_n\varepsilon)}. \tag{30}$$

These powers are negligible compared to $v_n^{-\gamma}$. In the case $\varepsilon=0$ (so that $\gamma>2$), set $\beta=0$ in Lemma 4 to obtain

$$\bar{E}_\sigma\left(\int_D \left|\sum_{j=j_n}^{J_n}\sum_k h_{jk}^{(2)}\psi_{jk}(x)\right|^\gamma \mathrm{d}x\right) \leqslant C(v,D,\gamma)\frac{(J_n-j_n)^{\gamma/2-1}}{n^{(\gamma-p)/2}}\sum_{j=j_n}^{J_n} j^{-p/2}$$

$$\leqslant C(v,D,\gamma)\left(\frac{J_n}{n}\right)^{(\gamma-p/2)}$$

where we used that for $\gamma>2$, $J_n/j_n \asymp \gamma/(\gamma-2)$. Thus this term has the right order.

(c4) Finally, we consider the important case $h_{jk}^{(4)}$. Let $\bar{\mathscr{B}}_j=\{k\in\mathscr{K}_j(\psi)\colon |d_{jk}|>2\lambda_n\}$. The condition $k\in\bar{\mathscr{B}}_j$ implies

$$|d_{jk}|\leqslant 2\lambda_n^{1/2}=\delta_n$$

say. In Donoho and Johnstone (1995) the following *modulus of continuity* was studied

$$\Omega^0(\delta;\|.\|,B)=\sup\{\|d\|\colon d\in B,\ |d_{jk}|\leqslant\delta,\ \forall j,k\}$$

where B is a (double indexed) sequence space and $\|\cdot\|$ is a norm on B. Using the embedding $B_{0\gamma\gamma\wedge 2}(\mathbb{R})\subset L_\gamma(\mathbb{R})$ and the characterization of Besov spaces in terms of sequence spaces, we have

$$\left(\int_D \left|\sum_{j=j_n}^{J_n}\sum_k h_{jk}^{(4)}\psi_{jk}(x)\right|^\gamma \mathrm{d}x\right)^{1/\gamma}\leqslant\|(\{d_{jk},\ j_n\leqslant j\leqslant J_n,\ k\in\bar{\mathscr{B}}_j\})\|_{0p'p'\wedge 2}$$

where $\|\cdot\|_{0p'p'\wedge 2}$ denotes the equivalent Besov norm on sequence spaces. It follows that

$$\left(\bar{E}_\sigma\left\{\int_D \left|\sum_{j=j_n}^{J_n}\sum_k h_{jk}^{(1)}\psi_{jk}(x)\right|^\gamma \mathrm{d}x\right\}\right)^{1/\gamma}\leqslant\Omega^0(\delta_n;\|.\|_{0p'p'\wedge 2},B_{spq})=\Omega_n,$$

say. From Theorem 3 in Donoho and Johnstone (1995), we have

$$\Omega_n\leqslant C(s,p,q,L,D)(\log n)^{e_C}\left(\frac{\log n}{n}\right)^\alpha \tag{31}$$

where

$$e_C=\begin{cases}0 & \varepsilon\neq 0,\\ \left(\frac{1}{2}-\frac{p}{p'q}\right)_+ & \varepsilon=0.\end{cases}$$

This bound is sharp when $\varepsilon\leqslant 0$. In the case $\varepsilon>0$, the exponent of $\log n$ can be improved to $(1-\varepsilon/sp)\alpha$ by a more detailed examination of the proof of Theorem 3 of Donoho and Johnstone (1995). The proof of Theorem 1 is complete.

2.2. Proof of Theorem 2

We follow classical methods for nonparametric lower bounds (see Korostelev and Tsybakov (1993) for instance). We evaluate lower bounds on a properly chosen parametric subfamily of functions in $B_{spq}\cap\mathscr{U}$.

We break the proof in two parts, the sparse case ($\varepsilon\leqslant 0$) and the dense case ($\varepsilon\geqslant 0$). The term "dense" and "sparse" refers to the number of basis functions used to form the perturbation. Without loss of generality, we restrict our attention to the case $b=0$ (the drift component) and assume that $D=[0,1]$.

2.2.1. The dense case

Given $g\in B_{spq}\cap\mathscr{U}$, we consider the family

$$\mathscr{C}_{j_n}=\left\{g+\rho_n\sum_{k\in K_{j_n}}\varepsilon_k\psi_{j_nk},\ \ \varepsilon_k=\pm 1\right\}\tag{32}$$

where $j_n\geqslant 0$ and $\rho_n>0$. The function ψ is a Daubechies wavelet with N vanishing moments, and we take $N\geqslant\lfloor s\rfloor+1$. Thus ψ has compact support in $[-A,A]$, where $A=2N-1$, and

$$K_{j_n}=\{-(2^{-j_n}-1)A+2lA,\ \ l=0,\dots,2^{j_n}-1\}$$

so that $\#\mathscr{C}_{j_n}=2^{2^{j_n}}$ which increases as n grows to infinity. Note that the functions ψ_{j_nk} and $\psi_{j_nk'}$ have disjoint support for $k\neq k'$.

For $\varepsilon\in\{-1,1\}^{K_{j_n}}$ we denote by $g^{\varepsilon}_{j_n}$, $\varepsilon=(\varepsilon_k,\ k\in K_{j_n})$ a generic element of $\mathscr{C}_{j_n}$ and by $P_{g^{\varepsilon}_{j_n}}$ the law of the observation $X^{(n)}=(X_{i/n},\ i=1,\dots,n)$ starting at time $1/n$ and specified by the parameter $g^{\varepsilon}_{j_n}$. Thus the vector $X^{(n)}$ has a density w.r.t. the Lebesgue measure on $\mathbb{R}^n$. Set

$$\Lambda(g^{\varepsilon}_{j_n},g^{\varepsilon'}_{j_n},X^{(n)})=\frac{\mathrm{d}P_{g^{\varepsilon}_{j_n}}}{\mathrm{d}P_{g^{\varepsilon'}_{j_n}}}(X^{(n)})$$

for the likelihood ratio between $P_{g^{\varepsilon}_{j_n}}$ and $P_{g^{\varepsilon'}_{j_n}}$. For fixed $k\in K_{j_n}$, define

$$g^{+}_{j_nk}=g+\rho_n\sum_{k'\neq k}\varepsilon_{k'}\psi_{j_nk'}+\gamma_n\psi_{j_nk},$$

$$g^{-}_{j_nk}=g+\rho_n\sum_{k'\neq k}\varepsilon_{k'}\psi_{j_nk'}-\gamma\psi_{j_nk}.$$

The following lemma can be found in Korostelev and Tsybakov (1993) in a slightly different version.

Lemma 5. *Let $\gamma \geqslant 1$. Assume that the following conditions are fullfilled*:

(i) $\forall \varepsilon \in \{-1,+1\}^{K_{j_n}}$, $g^{\varepsilon}_{j_n} \in B_{spq} \cap \mathscr{U}$.

(ii) *For large enough n, we have*

$$P_{g^-_{j_nk}}(\Lambda(g^+_{j_nk}, g^-_{j_nk}, X^{(n)}) > e^{-\lambda} \mid \mathscr{A}(v,D)) \geqslant p_0 > 0$$

where $\lambda > 0$ and p_0 are independent of n.

Then, for any estimator $\hat{g}_n$ and for large enough n

$$\max_{\varepsilon \in \{-1,+1\}^{K_{j_n}}} E_{g^{\varepsilon}_{j_nk}} \left\{ \int_D |\hat{g}_n(x) - g^{\varepsilon}_{j_nk}(x)|^{\gamma} \, \mathrm{d}x \mid \mathscr{A}(v,D) \right\}$$
$$\geqslant \frac{1}{2} p_0 \kappa(v,D) 2^{j_n \gamma/2} \int |\psi|^{\gamma} e^{-\lambda} \rho_n^{\gamma}$$

where $\kappa(v,D) = \inf_{\sigma^2 \in \mathscr{U}} P_{\sigma}(\inf_{x \in D_\delta} L_1^x \geqslant v)$.

Proof of Theorem 2 (*Dense case*). We take

$$g(x) = (m_2 - m_1)/2, \quad j_n = \left\lfloor \left(\frac{1}{1+2s} \right) \log_2 n \right\rfloor \quad \text{and} \quad \rho_n = \frac{m_0 2^{-j_n}}{\|\psi\|_\infty} \wedge \frac{m_1}{2} 2^{-j_n(s+1/2)}$$

so that condition (i) of Lemma 5 is satisfied, by simple calculation of its Besov norm using the wavelet expansion of $g^{\varepsilon}_{j_nk}$ (see Appendix). To prove (ii), we readily follow the proof of Proposition 1 in Hoffmann (1996). We conclude by applying Lemma 5 and noting that the factor $w_n \rho_n$ is asymptotically bounded below.

2.2.2. The sparse case

We now consider the family

$$\mathscr{P}_{j_n} = \{g, \; g_{j_nk} = g + \rho_n \psi_{j_nk}, \; k \in K_{j_n}\} \quad \text{for } j \geqslant 0. \tag{33}$$

Denote by $P_{g_{j_nk}}$ (resp. P_g) the law of the observations defined by the parameter g_{j_nk} (resp. P_g). Set

$$\Lambda(g, g_{j_nk}, X^{(n)}) = \frac{\mathrm{d}P_{g_{j_nk}}}{\mathrm{d}P_g}(X^{(n)}) \tag{34}$$

for the likelihood ratio between the parameters g_{j_nk} and g. The following result is again to be found in Korostelev and Tsybakov (1993) in a slightly different version.

Lemma 6. *Assume the following conditions are fulfilled*:

(i) $\forall k \in K_{j_n}$: $g_{j_nk} \in B_{spq} \cap \mathscr{U}$.

(ii) *The likelihood ratio* (34) *has the representation*

$$\Lambda(g, g_{j_nk}, X^{(n)}) = \exp(L_k^{(n)} - \log 2^{j_n} M_k^{(n)})$$

where $L_k^{(n)}$ and $M_k^{(n)}$ are random variables such that for large enough n

$$P_{g_{j_nk}}(L_k^{(n)} \geqslant 0) \geqslant p_0 > 0 \tag{35}$$

and

$$E_{g_{j_nk}}(|M_k^{(n)}|) \leqslant \frac{1}{2}\mu p_0 \tag{36}$$

for some p_0 *and* $0 \leqslant \mu < 1$ *independent of* n.

Then, for any estimator $\hat{g}_n$ *and large enough* n

$$\max_{k\in K_{j_n}} E_{g_{j_nk}}\left\{\int_D |\hat{g}_n(x) - g_{jk}(x)|^\gamma \mid \mathscr{A}(v,D)\right\}$$
$$\geqslant \kappa(v,D)\ \frac{1}{2}p_0\int |\psi|^\gamma(1-\mu)2^{j_n(\gamma/2-1)}\rho_n^\gamma.$$

Proof of Theorem 2 (*Sparse case*).

(a) We again take

$$g(x) = (m_2 - m_1)/2, \quad j_n = \left\lfloor\left(\frac{1}{1+2s-2/p}\right)\log_2 n\right\rfloor$$

and

$$\rho_n = L_0\frac{m_0 2^{-j_n}}{\|\psi\|_\infty} \wedge \frac{m_1}{2}2^{-j_n(s+1/2-1/p)}$$

where $L_0 > 0$ is a constant to be determined later so that condition (i) of Lemma 6 is satisfied.

(b) To prove (ii), we approximate the likelihood function by a first-order Gaussian transition, denoted by $\tilde{\Lambda}$, obtained when replacing the transition density of $(X_t, 0 \leqslant t \leqslant 1)$ by

$$\tilde{p}_\delta(x,y) = \frac{1}{\sqrt{2\pi\delta}}\frac{1}{\sigma(x)}\exp -\frac{1}{2\delta}\frac{(y-x)^2}{\sigma^2(x)}$$

while $X^{(n)}$ is still taken under $P_{g_{j_nk}}$. Under $P_{g_{j_nk}}$, we have

$$\log\tilde{\Lambda}(g, g_{j_nk}, X^{(n)})$$
$$= \sum_{i=1}^{n-1}\log\frac{1}{1+\gamma_n\psi_{j_nk}(X_{i/n})}$$
$$-\frac{1}{2}\sum_{i=0}^{n-1}\left[\left(\frac{1}{1+\gamma_n\psi_{j_nk}(X_{i/n})}\right)^2 - 1\right]n(W_{(i+1)/n} - W_{i/n})^2 + R_{1,k}^{(n)}. \tag{37}$$

Elementary computation leads to

$$\log\tilde{\Lambda}(g, g_{j_nk}, X^{(n)}) = -\rho_n\sum_{i=1}^{n-1}\psi_{j_nk}(X_{i/n})(n(W_{(i+1)/n} - W_{i/n})^2 - 1)$$
$$+\frac{1}{2}\rho_n^2\sum_{i=1}^{n-1}\psi_{j_nk}^2(X_{i/n})$$
$$-\frac{3}{2}\rho_n^2\sum_{i=1}^{n-1}\psi_{j_nk}^2(X_{i/n})(W_{(i+1)/n} - W_{i/n})^2 + R_{2,k}^{(n)}.$$

Set

$$\varepsilon_i = 1 - n(W_{(i+1)/n} - W_{i/n})^2$$

and define

$$L_k^{(n)} = -\rho_n \sum_{i=0}^{n-1} \psi_{j_n k}(X_{i/n})\varepsilon_i$$

$$M_k^{(n)} = -\frac{1}{2}\rho_n^2(\log 2^{j_n})^{-1} \sum_{i=1}^{n-1} \psi_{j_n k}^2(X_{i/n})$$

$$+ \frac{3}{2}\rho_n^2(\log 2^{j_n})^{-1} \sum_{i=1}^{n-1} \psi_{j_n k}^2(X_{i/n})(W_{(i+1)/n} - W_{i/n})^2 + R_k^{(n)}$$

where $R_k^{(n)}$ is a suitably renormalized remainder term. Thus we have the representation of (ii) for $\tilde{\Lambda}$. It remains to prove Eqs. (35) and (36). The arguments for replacing $\tilde{\Lambda}$ by Λ are obtained is the same line as in Proposition 1 of Hoffmann (1996) so we omit them.

(c) We first check that $P_{g_{j_n k}}(L_k^{(n)} \geqslant 0) \geqslant p_0$ for some $p_0 > 0$, or equivalently

$$P_{g_{j_n k}}\left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \psi_{j_n k}(X_{i/n})\varepsilon_i \geqslant 0\right) \geqslant p_0 > 0. \tag{38}$$

We will use the following elementary lemma.

Lemma 7. *Let* $(U_n,\ n \geqslant 0)$ *be a sequence of random variables such that* $E(U_n) = 0$, $0 < p_1 \leqslant E(U_n^2)$ *for large enough* n *and* $(U_n^2,\ n \geqslant 0)$ *is uniformly integrable. Then, there exists* $p_1 > 0$ *such that, for* n *large enough*

$$P(U_n \geqslant 0) \geqslant p_0 > 0. \tag{39}$$

Proof. Assume on the contrary that Eq. (39) is false. Then, up to taking a subsequence, we may assume that $P(U_n \geqslant 0) \to 0$. From the assumption, $E(|U_n| 1_{U_n \geqslant 0})$ is arbitrarily small for n large enough, and since $E(U_n) = 0$, $E(|U_n| 1_{U_n \leqslant 0})$ can be taken arbitrarily small as well. It follows that $E(|U_n|) \to 0$. Moreover

$$E(U_n^2) \leqslant E(|U_n| 1_{|U_n| \leqslant 1}) + E(|U_n^2| 1_{|U_n| \geqslant 1}). \tag{40}$$

The first term in the right-hand side of Eq. (40) converges to 0. We also have that $P(|U_n| \geqslant 1)$ converges to 0. Hence, from the uniform integrability of U_n^2, the second term in the right-hand side of Eq. (40) is arbitrarily small for large enough n. Hence $E(U_n^2)$ converges to 0, which contradicts $E(U_n^2) \geqslant p_1 > 0$.

We apply Lemma 7 to $U_n = (1/\sqrt{n}) \sum_{i=1}^{n} \psi_{j_n k}(X_{i/n})\varepsilon_i$. Clearly, $E_{g_{j_n k}}(U_n) = 0$. To show that $E_{g_{j_n k}}(U_n^2)$ is bounded below by some $p_1 > 0$ for large enough n, we will use an auxiliary result.

Lemma 8. *We have*

$$\limsup_{n\to\infty} \sup_{k \in K_{j_n}} \sup_{z \in \mathscr{S}(\psi_{j_n k})} E_{g_{j_n k}}(|U_n^2 - L_1^z|) = 0.$$

Proof. We write $U_n^2 - L_1^z = F_1^{(n)} + F_2^{(n)} + F_3^{(n)}$, where

$$F_1^{(n)} = \frac{1}{n}\sum_{i=1}^{n-1} \psi_{j_nk}^2(X_{i/n}) - L_1^z,$$

$$F_2^{(n)} = \frac{1}{n}\sum_{i=1}^{n-1} \psi_{j_nk}^2(X_{i/n})(\varepsilon_i^2 - 1),$$

$$F_3^{(n)} = \frac{1}{n^2}\sum_{i<j} \psi_{j_nk}^2(X_{i/n})\psi_{j_nk}^2(X_{j/n})\varepsilon_i\varepsilon_j.$$

First, note that $\psi_{j_nk}^2(x) = 2^{j_n}h(2^{j_n}x)$, where $h(x) = \psi^2(x-k)$. Hence the first term $F_1^{(n)}$ is a kernel approximation of the local time around $\mathscr{S}(\psi_{j_nk})$ and the convergence of $F_1^{(n)}$ follows. Clearly, $F_2^{(n)}$ and $F_3^{(n)}$ are zero-mean, and using that $\|\psi_{j_nk}\|_\infty \leqslant 2^{j_n/2}\|\psi\|_\infty$, it is easily seen that their variances both tend to zero uniformly in k.

Since $E_{g_{j_n}}(\inf_{x\in D} L_1^x) \geqslant v\kappa(v,D)$, the lower bound on $E_{g_{j_nk}}(U_n^2)$ follows by taking for instance $p_1 = v\kappa(v,D) - \eta$, for small enough $\eta > 0$. Likewise, the uniform integrability of U_n^2 is easily checked and Eq. (38) follows.

(d) Let us now turn to $M_k^{(n)}$. Note that $\rho_n \leqslant C_0 L_0\sqrt{\log n/n}$, where $C_0 = C_0(s,p,q)$. Clearly

$$E_{g_{j_nk}}(|M_k^{(n)}|) \leqslant L_0^2 C_0^2 \frac{2}{n}\sum_{i=0}^{n} E_{g_{j_nk}}(\psi_{j_nk}^2) + E_{g_{j_nk}}(|R_k^{(n)}|). \tag{41}$$

Using the rough inequality for the density $p_{i/n}(x_0,\cdot)$ of $X_{i/n}$

$$p_{i/n}(x_0,x) \leqslant C\sqrt{n/i}$$

where C only depends on D and m_i, $i=1,2,3$, we see that the first term in the right-hand side of Eq. (41) is bounded. Likewise, one easily check that $E_{g_{j_nk}}(|R_k^{(n)}|)$ converges to 0 uniformly in k. By taking L_0 small enough, the existence of $\mu < 1$ provided Eq. (36) follows.

Putting together (c) and (d), the assumptions of Lemma 6 hold and the conclusion of Theorem 2 in the sparse case follows. This ends the proof of Theorem 2.

2.3. *Proof of Theorem 3*

It is easily seen that under the specification of the levels j_n^* and J_n^*, Lemmas 1 and 2 remain valid. Define the indices $j_l(s,p,q)$, $l=1,0$, by

$$2^{j_0(s,p,q)} = (n(\log n)^{-I\{\varepsilon>0\}})^{1-2\alpha} \quad \text{and} \quad 2^{j_1(s,p,q)} = (n(\log n)^{-I\{\varepsilon\leqslant 0\}})^{\alpha/\bar{s}}$$

where $\bar{s} = s - 1/p + 1/\gamma$. Clearly $j_n^* \leqslant j_0(s,p,q) \leqslant j_n \leqslant j_1(s,p,q) \leqslant J_n^*$. We again use the decomposition of $\sigma^2(x) - T_n(x)$ used in the proof of Theorem 1.

(a) The terms $\mathscr{R}_n^{(1)}$ and $\mathscr{R}_n^{(2)}$ have rates of convergence no worse than in Theorem 1. More specifically

$$\mathscr{R}_n^{(2)} \leqslant C(v,D,\gamma) z_n^{-\gamma} 2^{-J_n^* \bar{s}\gamma} \leqslant C(v,D,\gamma) z_n^{-\gamma} 2^{-j_1(s,p,q)\bar{s}\gamma},$$

$$\mathscr{R}_n^{(1)} \leqslant C(v,D,\gamma)\left(z_n^{-\gamma}\frac{2^{j_n^*}}{n}\right)^{\gamma/2} \leqslant C(v,D,\gamma) z_n^{-\gamma}\left(\frac{2^{j_n}}{n}\right)^{\gamma/2}.$$

(b) The asymptotic behaviour of the terms of (c1) and (c2) is treated exactly as in Theorem 1.

(c) The proof for the term in (c4) remains unchanged. We focus on the term in (c3). We first look at the case $\varepsilon \leqslant 0$ which implies $\gamma > 2$. Let $p < p_1 < \gamma$ be defined by $(\gamma - p_1)/2 = \alpha\gamma$. Applying Lemma 4 for $\beta = 0$ gives

$$\sum_k \bar{E}_\sigma(|\hat{h}_{jk}^{(2)}\psi_{jk}|^\gamma) \leqslant C(v,D,\gamma)(J_n^* - j_n^*)^{(\gamma-2)/2} n^{\gamma/2} \sum_{j_n^*}^{J_n^*} 2^{j(\gamma/2-1)} \sum_k \left| 2 d_{jk} \lambda_n \sqrt{\frac{n}{j}} \right|^{p_1}$$

$$\leqslant C(v,D,\gamma)\left(\frac{J_n^*}{n}\right)^{(\gamma-p_1)/2} \left[\sup_j 2^{j(\gamma-2)/2p_1} \|d_{j.}\|_{p_1}\right].$$

Since $\varepsilon \leqslant 0$, we have $(\gamma - p)/2 \geqslant \alpha\gamma$. Set $\tilde{s} = s + 1/2 - 1/p$. It follows that

$$p_1 = \gamma(1 - \bar{s}/\tilde{s}) = (\gamma - 2)/2\tilde{s}.$$

Hence $(\gamma - 2)/2p_1 = \tilde{s}$, and since $\|d_{j.}\|_p$ increases as p decreases, the above supremum is bounded by $\|\sigma^2\|_{sp\infty}$. Thus

$$\sum_k \bar{E}_\sigma(|\hat{h}_{jk}^{(2)}|^\gamma) \leqslant C(v,D,\gamma)\left(\frac{J_n^*}{n}\right)^{\alpha\gamma} \|\sigma^2\|_{sp\infty}^{p_1} \leqslant C(v,D,\gamma)\left(\frac{\log n}{n}\right)^{\alpha\gamma}.$$

When $\varepsilon > 0$, we write

$$\sum_k \hat{h}_{jk}^{(2)}\psi_{jk}(x) = \left(\sum_{j_n^*}^{j_1(spq)} + \sum_{j_1(spq)}^{J_n^*}\right) \sum_k (\hat{d}_{jk} - d_{jk}) 1_{\{|\hat{d}_{jk}| \geqslant \lambda_{jn}, |d_{jk}| \geqslant \lambda_{jn}/2\}} \psi_{jk}(x)$$

$$= F_1^{(n)}(x) + F_2^{(n)}(x),$$

say. The term $F_2^{(n)}$ is bounded exactly like in the previous section since the upper limit J_n^* does not affect the estimate. For $F_1^{(n)}$, we exploit Lemma 4 along with Lemma 1 applied to $\hat{d}_{jk}$ instead of $\hat{c}_{jk}$ to obtain

$$\bar{E}_\sigma\left(\int |F_1^{(n)}(x)|^\gamma \,\mathrm{d}x\right) \leqslant C(v,D,\gamma) S(\beta a)^{(\gamma/2-1)_+} \sum_{j_n^*}^{j_1(s,p,q)} 2^{-j\beta\gamma/2}\left(\frac{2^j}{n}\right)^{\gamma/2}$$

$$\leqslant C(v,D,\gamma)\left(\frac{2^{j_1(s,p,q)}}{n}\right)^{\gamma/2} \leqslant C(v,D,\gamma) n^{-\alpha\gamma}.$$

The proof of Theorem 3 is complete.

3. Extension to other diffusion models

We show in this section how to obtain results similar to Theorems 1–3 in other diffusion models. The forthcoming results will be given without proof but the modifications from Theorems 1 and 3 which are needed will be sketched.

3.1. Time-dependent diffusion coefficient

We consider a 1-dimensional diffusion process X defined by

$$X_t = x_0 + \int_0^t b(s, X_s)\,\mathrm{d}s + \int_0^t \sigma(s)\,\mathrm{d}W_s, \tag{42}$$

where $x_0 \in \mathbb{R}$, W is a standard Brownian motion and b and σ are unknown. We assume A and

Assumption C. The function σ is nonvanishing and σ^2 is Lipschitz continuous over $[0,1]$.

Under Assumptions A and C, the process X is well defined. We aim at estimating $\sigma^2(t)$ from the observation $(X_{i/n},\ i=0,\dots,n)$. The decomposition (2) of Section 1 now becomes

$$n(X_{(i+1)/n} - X_{i/n})^2 = n \int_{i/n}^{(i+1)/n} \sigma^2(s)\,\mathrm{d}s + \varepsilon_{i/n} + \text{a term of higher order} \tag{43}$$

where $\varepsilon_{i/n} = (\int_{i/n}^{(i+1)/n} \sigma(s)\,\mathrm{d}W_s)^2 - \int_{(i-1)/n}^{i/n} \sigma^2(s)\,\mathrm{d}s$. The variables $\varepsilon_{i/n}$ are centred and independent. The regression approximation reads

$$n(X_{(i+1)/n} - X_{i/n})^2 \simeq \sigma^2(i/n) + \varepsilon_{i/n}.$$

Thus, we are in the most favourable case, with fixed design and independent noises. To estimate $\sigma^2(t)$, we simply estimate the wavelet coefficients by

$$\tilde{c}_{jk} = \sum_{i=0}^{n-1} \varphi_{jk}(i/n)(X_{(i+1)/n} - X_{i/n})^2 \quad \text{and} \quad \tilde{d}_{jk} = \sum_{i=0}^{n-1} \psi_{jk}(i/n)(X_{(i+1)/n} - X_{i/n})^2 \tag{44}$$

and we construct an estimator $\tilde{T}_n$ with formula (6) using $\tilde{c}_{jk}$ and $\tilde{d}_{jk}$. Given a compact subset D of $[0,1]$, $\gamma \in [1,\infty)$ and a set of constraint Σ we consider the following minimax risk:

$$\mathscr{R}_{D,\gamma}\{T_n, \alpha_n, \Sigma\} = \sup_{\sigma^2 \in \Sigma} E_\sigma \left\{ \alpha_n^{-\gamma} \int_D |\sigma^2(t) - T_n(t)|^\gamma\,\mathrm{d}t \right\}^{1/\gamma}$$

where $\alpha_n > 0$ is a normalizing factor.

Theorem 4. *Grant Assumptions* A, C *and let* $D \subset (0,1)$ *be a compact interval. Let* $s > 1 + 1/p$, $1 \leqslant p \leqslant \gamma < \infty$, $q \in [1,\infty]$.

Let T_n *be specified by* (j_n, J_n, λ_n) *from Eqs.* (8) *and* (9) *of Section* 1, *and let* $N \geqslant \lfloor s \rfloor + 1$.

(a) *We have* $\limsup_{n\to\infty} \mathscr{R}_{D,\gamma}\{T_n, v_n, B_{spq}([0,1], L)\} < +\infty$.

(b) *We have* $\liminf_{n\to\infty} \inf_{\hat{T}_n} \mathscr{R}_{D,\gamma}\{\hat{T}_n, w_n, B_{spq}([0,1], L)\} > 0$, *where the infimum is taken over all possible estimators.*

(c) *With the notations of Section* 1, $T_n^* = T_n(j_n^*, J_n^*, \lambda_n^*)$, *with* $(j_n^*, J_n^*, \lambda_n^*)$ *specified by Eqs.* (10) *and* (11) $N \geqslant r_0 + 1$ *is nearly adaptive over* $\mathscr{S}$:

$$\sup_{(s,p,q)\in\mathscr{S}} \limsup_{n\to\infty} \mathscr{R}_{D,\gamma}\{T_n, z_n, B_{spq}([0,1], L)\} < +\infty.$$

This results extend former results of Genon-Catalot et al. (1992).

3.2. Ergodic diffusions

We consider a 1-dimensional homogeneous diffusion process X solving, for $t \in [0, \infty)$

$$X_t = x_0 + \int_0^t b(X_s)\,\mathrm{d}s + \int_0^t \sigma(X_s)\,\mathrm{d}W_s, \tag{45}$$

where $x_0 \in \mathbb{R}$, W is a standard Brownian motion and b and σ are unknown. Let m_i, $i = 0, \dots, 5$ be given constants. The assumptions on b and σ are more stringent than in Section 1.

Assumption A′. The function b belongs to the set $\mathscr{V}$ defined by

$$\mathscr{V} = \{g \in \mathscr{C}^2(\mathbb{R}):\ \|g'\|_\infty \leqslant m_0,\ \|g''\|_\infty \leqslant m_1,\ \mathrm{sgn}(x)g(x) \leqslant m_2\}$$

with the notation $\mathrm{sgn}(x) = 1$ if $x \geqslant 0$ and -1 otherwise.

Assumption B′. The function σ^2 belongs to the set $\mathscr{W}$ defined by

$$\mathscr{W} = \{g \in \mathscr{C}^2(\mathbb{R}):\ m_3 \leqslant g(x) \leqslant m_4,\ \|g'\|_\infty \leqslant m_5,\ \|g''\|_\infty \leqslant m_6\}.$$

Under Assumptions A′ and B′, the process X is geometrically ergodic over $\mathbb{R}$ and has a unique invariant density $\mu_{\sigma,b}$. We will denote by $P_{\sigma,b}$ the law of X.

We aim at estimating both $b(x)$ and $\sigma^2(x)$ from the observation $(X_{i\Delta_n},\ i = 0, \dots, n)$, with $\Delta_n \to 0$ and $n\Delta_n \to +\infty$ as $n \to \infty$.

(1) To estimate the drift $b(x)$, we first note that

$$\Delta_n^{-1}(X_{(i+1)\Delta_n} - X_{i\Delta_n}) = \Delta_n^{-1} \int_{i\Delta_n}^{(i+1)\Delta_n} b(X_s)\,\mathrm{d}s + \varepsilon_{i/n}$$

with $\varepsilon_{i/n} = \Delta_n^{-1} \int_{i\Delta_n}^{(i+1)\Delta_n} \sigma(X_s)\,\mathrm{d}W_s$. Thus we are led again to a regression framework, since $\Delta_n^{-1} \int_{(i-1)\Delta_n}^{i\Delta_n} b(X_s)\,\mathrm{d}s \simeq b(X_{i\Delta_n})$ if Δ_n is small enough. The noise term $\varepsilon_{i/n}$ has a variance of order Δ_n^{-1} and explodes as $n \to \infty$. This will affect the rates of convergence for the estimation of $b(x)$.

Following the procedure described in Section 1, replacing the times i/n by $i\Delta_n$, we estimate the wavelet coefficients c_{jk} and d_{jk} of b by

$$\bar{c}_{jk} = \frac{1}{\lfloor nv \rfloor} \sum_{i=1}^{\lfloor nv \rfloor} \varphi_{jk}(\xi_{\tau_i}) \Delta_n^{-1} (X_{\tau_i + \Delta_n} - X_{\tau_i}) \tag{46}$$

and

$$\bar{d}_{jk} = \frac{1}{\lfloor nv \rfloor} \sum_{i=1}^{\lfloor nv \rfloor} \psi_{jk}(\xi_{\tau_i}) \Delta_n^{-1} (X_{\tau_i + \Delta_n} - X_{\tau_i}) \tag{47}$$

and we construct an estimator $\Lambda_n(x)$ of $b(x)$ using formula (6) of Section 1 along with the $\bar{c}_{jk}$ and $\bar{d}_{jk}$.

The requirement that enough observations are available around x to estimate $b(x)$ can be transferred into a condition on the invariant density $\mu_{\sigma,b}$ of X, namely that $\mu_{\sigma,b}$ is bounded away from zero in a neighbourhood of x, uniformly in (σ, b). We have

$$\mu_{\sigma,b}(x) = \frac{C_{\sigma,\mu}}{\sigma^2(x)} \exp\left(2 \int_0^x \frac{b(u)}{\sigma^2(u)} \, \mathrm{d}u \right)$$

where $C_{\sigma,\mu}$ is a normalizing constant. Straightforward computations show that if D is a compact subset of $\mathbb{R}$

$$\inf_{(\sigma^2,b)\in\mathscr{V}\times\mathscr{W}} \inf_{x\in D} \mu_{\sigma,b} = v > 0$$

where $v = v(D, m_i, 1 \leqslant i \leqslant 5)$ which suggests a choice for v in the construction of Λ_n. Finally, if $\alpha_n = \alpha(n)$ is a real-valued sequence, we will denote by α_n^Δ the sequence with values $\alpha(n\Delta_n)$. For a set of constraint Σ, we choose the risk

$$\mathscr{R}_{D,\gamma}^{\text{drift}}\{\Lambda_n, \alpha_n, \Sigma\} = \sup_{b\in\Sigma} E_{\sigma,b} \left\{ \alpha_n^{-\gamma} \int_D |\Lambda_n(x) - b(x)|^\gamma \, \mathrm{d}x \right\}^{1/\gamma}$$

for a normalizing factor $\alpha_n > 0$. The performances of Λ_n are summarized in the following theorem:

Theorem 5 (The drift function). *Grant Assumptions* A$'$, B$'$ *and let* $n\Delta_n \to \infty$, $n\Delta_n^2 \to 0$. *Let* $s > 2 + 1/p$, $1 \leqslant \gamma < \infty$, $q \in [1, \infty]$ *and* D *be a compact interval of* $\mathbb{R}$.

(a) *Let* Λ_n *be specified by*, $(j_n^\Delta, J_n^\Delta, \lambda_n^\Delta, h_n)$ *from Eqs.* (8) *and* (9), *together with* $v(D)$ *and* $N \geqslant \lfloor s \rfloor + 1$. *We have*

$$\limsup_{n\to\infty} \mathscr{R}_{D,\gamma}^{\text{drift}}\{\Lambda_n, v_n^\Delta, B_{spq} \cap \mathscr{V}\} < \infty.$$

(b) *The following lower bound holds*

$$\liminf_{n\to\infty} \inf_{\tilde{\Lambda}_n} \mathscr{R}_{D,\gamma}^{\text{drift}}\{\tilde{\Lambda}_n, w_n^\Delta, B_{spq} \cap \mathscr{V}\} > 0,$$

where the infimum is taken over all possible estimators.

(c) *The same adaptation results as in Theorem* 3 *can be derived for* Λ_n, *replacing* n *by* $n\Delta_n$.

(2) To estimate $\sigma^2(x)$, we use exactly the same procedure as in Section 1, taking into account the new observation scheme $(i\Delta_n,\ i=0,\dots,n)$. Formally, we obtain an estimator $M_n(x)$ of $\sigma^2(x)$ by replacing the time step $1/n$ by Δ_n in the construction of the estimator $T_n(x)$ of Section 1. We choose the risk

$$\mathscr{R}^{\mathrm{diff}}_{D,\gamma}\{M_n,\alpha_n,\Sigma\} = \sup_{\sigma^2\in\Sigma} E_{\sigma,b}\left\{\alpha_n^{-\gamma}\int_D |M_n(x)-\sigma^2(x)|^\gamma\,\mathrm{d}x\right\}^{1/\gamma}$$

and we obtain the following result.

Theorem 6 (The diffusion coefficient). *Grant Assumptions* A′ *and* B′ *and let* $n\Delta_n\to\infty$, $n\Delta_n^2\to 0$. *Let* M_n *be specified by* (j_n,J_n,λ_n,h_n) *from Eqs.* (8) *and* (9), *together with* $v(D)$ *and* $N\geqslant\lfloor s\rfloor+1$.

(a) *If* $s>2+1/p$, *Theorem* 1 *holds for* M_n, *replacing* $\mathscr{R}_{D,\gamma,v}$ *by* $\mathscr{R}^{\mathrm{diff}}_{D,\gamma}$ *and* $\mathscr{U}$ *by* $\mathscr{W}$.

(b) *If, in addition,* $s>3+1/p$, *we have Theorem* 2, *replacing* $\mathscr{R}_{D,\gamma,v}$ *by* $\mathscr{R}^{\mathrm{diff}}_{D,\gamma}$ *and* $\mathscr{U}$ *by* $\mathscr{W}$.

(c) *The same adaptation results as in Theorem* 3 *can be derived for* M_n.

Remark. (1) As we might expect, the rates for estimating $b(x)$ and $\sigma^2(x)$ differ.

(2) The smoothness of $\sigma^2(x)$ does not affect the rate of convergence for the estimation of $b(x)$ and vice versa.

Acknowledgements

I am grateful to Jean Jacod and Dominique Picard for helpful discussions and comments.

Appendix

We give some basic properties and definition of multiscale decomposition and Besov spaces, following Cohen (1998). More data on the subject may be found in classical textbooks (for instance Daubechies, 1992; Meyer, 1990).

A.1. Characterization of Besov spaces by wavelet sequences

One can construct a function φ (the *scaling function*) such that

(i) The sequence $\{\varphi(x-k),\ k\in\mathbb{Z}\}$ is an orthonormal family of $L_2(\mathbb{R})$.

(ii) If V_j denotes the subspace spanned by $\{\phi_{jk},\ k\in\mathbb{Z}\}$ then $\forall j\in\mathbb{Z},\ V_j\subset V_{j+1}$. We then have $\bigcap_{j\in\mathbb{Z}}V_j=\{0\}$ and $\bigcup_{j\in\mathbb{Z}}V_j$ is dense in $L_2(\mathbb{R})$.

(iii) φ is of class $\mathscr{C}^r$ and compactly supported. Under these conditions, define the space W_j by

$$V_{j+1}=V_j\oplus W_j.$$

Then, there exists a function ψ (called *the wavelet*) such that

(iv) $\{\psi(x-k),\ k\in\mathbb{Z}\}$ is an orthonormal basis of W_0.

(v) $\{\psi_{jk},\ k\in\mathbb{Z},\ j\in\mathbb{Z}\}$ is an orthonormal basis of $L_2(\mathbb{R})$, where $\psi_{jk}(x)=2^{j/2}\psi(2^jx-k)$.

(vi) ψ has the same regularity and localization properties as φ.

In addition, we have the following decomposition for any integer j_0:

$$L_2(\mathbb{R})=V_{j_0}\oplus\bigoplus_{j\geqslant j_0}W_j.$$

For $j_0\in\mathbb{Z}$, the following expansion holds for a function $f\in L_2(\mathbb{R})$:

$$f=\sum_{k\in\mathbb{Z}}c_{j_0k}\varphi_{j_0k}+\sum_{j\geqslant j_0}\sum_{k\in\mathbb{Z}}d_{jk}\psi_{jk}$$

where $c_{j_0k}=\int f(x)\varphi_{j_0k}(x)\,\mathrm{d}x$ and $d_{jk}=\int f(x)\psi_{jk}(x)\,\mathrm{d}x$.

Let P_j be the associated projection operator onto V_j and $D_j=P_{j+1}-P_j$. Besov spaces depend on three parameters: $s>0$ (regularity parameter), $1\leqslant p\leqslant\infty$ (L_p parameter) and $1\leqslant q\leqslant\infty$ (interpolation parameter) and are denoted by $B_{spq}(\mathbb{R})$.

Definition 1. A function f belongs to the space $B_{spq}(\mathbb{R})$ if and only if, for any j_0, the norm

$$J_{spq}(f)=\|P_{j_0}f\|_{L_p(\mathbb{R})}+\left(\sum_{j\geqslant j_0}(2^{js}\|D_jf\|_p)^q\right)^{1/q}<\infty$$

if $q<\infty$ and

$$J_{sp\infty}(f)=\|P_{j_0}f\|_{L_p(\mathbb{R})}+\sup_{j\geqslant j_0}(2^{js}\|D_jf\|_p)<\infty$$

otherwise.

Using now the decomposition of f

$$P_{j_0}f=\sum_{k\in\mathbb{Z}}c_{j_0k}\varphi_{j_0k}$$
$$D_jf=\sum_{k\in\mathbb{Z}}d_{jk}\psi_{jk}$$

we may also say that $f\in B_{spq}(\mathbb{R})$ if the equivalent norm

$$J'_{spq}(f)=\|c_{j_0.}\|_{l^p}+\left(\sum_{j\geqslant j_0}(2^{j(s+1/2-1/p)}\|d_{j.}\|_{l^p})^q\right)^{1/q}<\infty$$

where we have set $\|d_{j.}\|_{l^p}=(\sum_{k\in\mathbb{Z}}|\beta_{j,k}|^p)^{1/p}$. This second definition is equivalent to the first one as a consequence of Meyer's lemma (see Meyer, 1990, p. 30).

Well known particular cases of Besov spaces include the Sobolev spaces $H^s(\mathbb{R})=B_{s22}(\mathbb{R})$ and the set of Hölder functions $B_{s\infty\infty}(\mathbb{R})$ when s is not an integer.

A.2. Multiscale decompositions and thresholding

Multiscale methods are based on approximations of the data of a given problem at various (here dyadic) resolution levels. A *multiscale decomposition* is obtained by expanding a function f into the sum of a coarse approximation at a level j_0 and additional details

$$f = f_{j_0} + \sum_{j \geqslant j_0} g_j \tag{48}$$

where each $g_j = f_{j+1} - f_j$ represents the fluctuations of f between the two successive levels j and $j+1$. The coarse approximation f_{j_0} at level j_0 is a sketchy picture of f which does not oscillate at a frequency higher than 2^{j_0}. The existence of *orthonormal wavelet bases* allows to have a decomposition (48) of the form

$$g_j = \sum_{k\in\mathbb{Z}} d_{jk}\psi_{jk} \tag{49}$$

where $d_{jk} = \int f(x)\psi_{jk}(x)\,\mathrm{d}x$ and $\psi_{jk}(x) = 2^{j/2}\psi(2^j x - k)$. The wavelet ψ is well localized both in time and frequency, and *oscillates*, in the sense that

$$\int x^l \psi(x)\,\mathrm{d}x = 0, \quad l = 0,\ldots,N \tag{50}$$

for some integer N. Each wavelet ψ_{jk} contributes to the fluctuation of f at scale 2^{-j} in a neighbourhood of size $2^{-j}|\mathrm{Supp}(\psi)|$ around $k2^{-j}$. It can be shown that f_{j_0} can also be written in the form

$$f_{j_0} = \sum_{k\in\mathbb{Z}} c_{j_0k}\varphi_{j_0k}$$

where φ is a scaling function, $c_{j_0k} = \int f(x)\varphi_{j_0k}(x)\,\mathrm{d}x$ and $\varphi_{j_0k}(x) = 2^{j_0/2}\varphi(2^{j_0}x - k)$.

The advantage of the multiscale structure follows from the cancellation property (50) of the wavelet ψ: it implies that d_{jk} is significantly small if f is smooth around $k2^{-j}$. For instance, if f is m times differentiable around $k2^{-j}$, then $|d_{jk}| \leqslant C2^{-j(m+1/2)}$. This has crucial consequences for estimating a smooth function f from noisy data: intuitively, the "small" wavelet coefficients estimates contain only "noise" and the *nonlinear* procedure

$$\sum_{k\in\mathbb{Z}} \hat{c}_{j_0k}\varphi_{j_0k} + \sum_{j_0}^{j_1}\sum_{k\in\mathbb{Z}} \hat{d}_{jk} 1_{\{|\hat{d}_{jk}| \geqslant \lambda\}}\psi_{jk}$$

where $\hat{c}_{j_0k}$ and $\hat{d}_{jk}$ are estimates of c_{j_0k} and d_{jk} will essentially be better than its linear analog

$$\sum_{k\in\mathbb{Z}} \hat{c}_{j_1}\varphi_{j_1k}$$

for a proper choice of φ, ψ, j_0, j_1 and λ.

References

Barlow, M.T., Yor, M., 1982. Semi-martingale inequalities via the Garsia-Rodemich-Rumsey lemma and application to local times. J. Funct. Anal. 49, 198–229.

Cohen, A., 1998. Multiscale methods and wavelets in numerical analysis. Handbook of Numer. Anal., to appear.

Daubechies, I., 1992. Ten lectures on wavelets, 'CBMS-NSF Series in Applied Mathematics', vol. 61, SIAM, Philadelphia.

Donoho, D.L., Johnstone, I.M., 1995. Minimax estimation via wavelet shrinkage, Ann. Statist., to appear.

Donoho, D.L., Johnstone, I.M., Kerkyacharian, G., Picard, D., 1995. Wavelet shrinkage, Asymptopia ? J. Roy. Statist. Soc. Ser. B 57, 301–369. With discussion.

Donoho, D.L., Johnstone, I.M., Kerkyacharian, G., Picard, D., 1996. Density estimation by wavelet thresholding. Ann. Statist. 24, 508–539.

Florens-Zmirou, D., 1993. On estimating the diffusion coefficient from discrete observations. J. Appl. Probab. 30, 790–804.

Genon-Catalot, V., Laredo, C., Picard, D., 1992. Nonparametric estimation of the diffusion coefficient by wavelet methods. Scand. J. Statist. 19, 319–335.

Hall, P., Heyde, C.C., 1980. Martingale Limit Theory and Its Applications. Academic Press, New York.

Hoffmann, M., 1996. L_p estimation of the diffusion coefficient. Bernoulli, to appear.

Jacod, J., 1997. Nonparametric kernel estimation of the diffusion coefficient of a diffusion. Preprint 405, Laboratoire de Probabilités, Université Paris 6.

Korostelev, A.P., Tsybakov, A.D., 1993. Minimax Theory of Image Reconstruction. Lecture Notes in Statistics, vol. 82. Springer, Berlin.

Kerkyacharian, G., Picard, D., 1993. Density estimation by kernel and wavelets methodes: optimality of Besov spaces. Statist. Probab. Lett. 18, 327–336.

Lipster, R.S., Shiryayev, A.N., 1977. Statistics of random processes I: General Theory. Springer, Berlin.

Meyer, Y., 1990. Ondelettes et Opérateurs I. Hermann, Paris.