



**HAL**  
open science

# Efficient big data analysis for ergodic diffusion models on the basis of discrete data

L I Galtchouk, S M Pergamenshchikov

► **To cite this version:**

L I Galtchouk, S M Pergamenshchikov. Efficient big data analysis for ergodic diffusion models on the basis of discrete data. 2020. hal-02474675

**HAL Id: hal-02474675**

**<https://hal.science/hal-02474675>**

Preprint submitted on 11 Feb 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Efficient big data analysis for ergodic diffusion models on the basis of discrete data. \*

L.I. Galtchouk <sup>†</sup>      S.M. Pergamenschikov<sup>‡</sup>

## Abstract

We consider a big data analysis problem for diffusion processes in the framework of nonparametric estimation for ergodic diffusion processes based on observations at discrete time moments in the case when diffusion coefficients are unknown. To this end we use the model selection method developed in [21]. In this paper through the oracle inequalities from [21] we show that the constructed model selection procedures are asymptotically efficient in adaptive setting, i.e. in the case when the regularity of the drift coefficient is unknown. To this end, for the first time for such problem, we found in the explicit form the celebrated Pinsker constant which is the sharp lower bound for the minimax squared accuracy normalized by the optimal convergence rate, i.e. it provides the best potentially possible estimation accuracy. Finally, we show that the accuracy of the model selection procedure asymptotically coincides with this lower bound, i.e. the constructed procedure is efficient.

*Keywords:* Adaptive nonparametric drift estimation; Asymptotic efficiency; Discrete time data; Nonasymptotic estimation; Model selection; Quadratic risk; Sharp oracle inequality.

*AMS 2000 Subject Classifications:* primary 62G08; secondary 62G05, 62G20.

---

\*This paper is supported by the Normandy RIN projects FuMa and MaSyComB

<sup>†</sup>International Laboratory of Statistics of Stochastic Processes and Quantitative Finance, Tomsk State University, 36, Lenin str., 634041, Tomsk, Russia; e-mail: leonid.galtchouk0667@orange.fr

<sup>‡</sup>Laboratoire de Mathématiques Raphael Salem, Avenue de l'Université, BP. 12, Université de Rouen, F76801, Saint Etienne du Rouvray, Cedex France and Department of Mathematics and Mechanics, Tomsk State University, Lenin str. 36, 634041 Tomsk, Russia; e-mail: Serge.Pergamenschikov@univ-rouen.fr

# 1 Introduction

## 1.1 Problem

In this paper we deal with the process  $(y_t)_{t \geq 0}$  defined by the stochastic differential equation

$$dy_t = S(y_t) dt + b(y_t) dW_t, \quad 0 \leq t \leq T, \quad (1.1)$$

where  $(W_t)_{t \geq 0}$  is a standard Wiener process, the initial value  $y_0$  is a fixed constant,  $S(\cdot)$  is unknown drift and  $b(\cdot)$  is unknown diffusion coefficient. The problem is to estimate the function  $S(x)$ , for  $x \in [\mathbf{x}_0, \mathbf{x}_1]$ , on the basis of the observations

$$(y_{t_j})_{1 \leq j \leq N}, \quad t_j = j\delta, \quad (1.2)$$

where the frequency  $\delta = \delta_T \in (0, 1)$  and the sample size  $N = N(T)$  are some functions of  $T$  that will be specified later. We consider the quadratic risk defined, for any estimator  $\widehat{S}$ , as

$$\mathcal{R}_\vartheta(\widehat{S}) = \mathbf{E}_\vartheta \|\widehat{S} - S\|^2 \quad \text{and} \quad \|f\|^2 = \int_{\mathbf{x}_0}^{\mathbf{x}_1} |f(x)|^2 dx, \quad (1.3)$$

where  $\mathbf{E}_\vartheta$  is the expectation with respect to the distribution of the process (1.1) for the functions  $\vartheta = \vartheta(\cdot) = (S(\cdot), b(\cdot))$ .

## 1.2 Motivations

The main motivation for the nonparametric drift in (1.1) is given in the papers [7, 8, 21] in which Big Data problems are studied in the framework of the high dimension linear diffusion models, i.e. the model (1.1) with

$$S(x) = \sum_{j=1}^q \theta_j \psi_j(x), \quad (1.4)$$

where  $(\psi_j)_{1 \leq j \leq q}$  are known linear independent functions and the dimension is greater than the number of observations, i.e.  $q > N$ . Indeed, such problems are important at various applications such that stochastic optimal control, finance, filtration, signal processing and etc (see, for example, [1, 26, 32, 29, 28, 2] and the references therein). Nonparametric estimation problems for  $S$  were studied in a number of papers in the case of complete observations, that is when a whole trajectory  $(y_t)_{0 \leq t \leq T}$  was observed. A sufficiently complete survey one can find in [30]. It should be noted that, for

the first time, the famous Pinsker constant representing the asymptotic efficiency property for nonparametric diffusion models was found by Dalayan and Kutoyants in [4, 5] for a special weighted integral risk using very nice local time tool. In non asymptotic setting Galtchouk and Pergamenschikov in [10, 11, 12, 13] developed nonparametric sequential estimation methods for the models (1.1) on the basis of which in [17] they calculated the Pinsker constant for the risk (1.3). It should be noted that in all the cited papers estimation problems were studied in the case of complete observations  $(y_t)_{0 \leq t \leq T}$ . In practice, usually for the models (1.1) the observations are accessible only at the discrete time moments (1.2). A natural question arises about properties of estimators based on discrete observations. The nonparametric estimation based on discrete time observations for models of the form (1.1) was considered firstly for estimating the unknown diffusion coefficient  $b(\cdot)$  on a fixed interval  $[0, T]$ , when the observation frequency goes to zero, (see, for example, [6], [24], [25], [27] and the references therein). Later, in [23] kernel estimates of the drift and diffusion coefficients were studied for the reflected processes (1.1) with the values in the interval  $[0, 1]$ . Minimax optimal convergence rates are found as the sample size goes to infinity. As to the estimation in the ergodic case, it should be noted that firstly a sequential procedure was proposed in [25] for nonparametric estimating the drift coefficient of the process (1.1) in the integral metric. Some upper and lower asymptotic bounds were found for the  $\mathbf{L}_p$  - risks. Later, in the paper [3] a non-asymptotic oracle inequality was obtained for a special empiric quadratic risk defined as a function of the observations at the discrete time moments. In the asymptotic setting, when the observation frequency goes to zero and the length of the observation time interval tends to infinity, the constructed estimators reach the minimax optimal convergence rates. But, unfortunately, neither the optimal convergence rate and nor efficiency property were established in [3].

### 1.3 Key ideas

Our approach is based on the sequential analysis methods developed in the papers [10, 12, 13] for nonparametric estimation problems. This approach makes possible to replace the random denominator by a conditional constant in a sequential Nadaraya-Watson estimator. Let us recall that in the case of complete observations (that is, when a whole trajectory is observed) the sequential estimator efficiency was proved by making use of a uniform concentration inequality (see [14]), besides an indicator kernel estimator. As it turns out later in [17], the efficient kernel estimate in the above given sense provides constructing a selection model adaptive procedure that appears ef-

efficient in the quadratic metric. Therefore, in order to realize this program (i.e. from efficient pointwise estimators to an efficient  $\mathbf{L}_2$ -estimator) in the case of discrete time observations, one needs to use a suitable concentration inequalities. Such concentration inequalities are obtained in [18] through the uniform geometric ergodicity method for the process (1.1) developed in [19] which provides non asymptotic upper bounds uniformly over functions  $S(\cdot)$  and  $b(\cdot)$  for the convergence rate in the ergodic theorem. Using this tool we can show that the corresponding weight least square estimator for  $S$  in (1.1) setting on the regularity parameters is efficient for the risk (1.3) in the Pinsker sense [33]. Finally, using the sharp oracle inequalities obtained in [21] we can estimate asymptotically from above the risk of the model selection procedure with the risk efficient weight least square estimator and we obtain the efficiency property in the adaptive sense, i.e. don't using the regularity of the function  $S$ .

## 1.4 Plan of the paper

The paper is organized as follows. In Section 2 we describe the functional classes. In Section 3 a sequential estimator of the drift coefficient is constructed. In Section 4 we introduce a regression model based on the above sequential estimator. In Section 5 we construct the model selection procedure and announce the corresponding oracle inequalities. The main results on asymptotic efficiency are formulated in Section 6. In Section 7 we study main asymptotic properties of the basic regression model. In Section 8 we study the efficiency property for non adaptive setting, i.e. in the case when the regularity of  $S$  and the Pinsker variance are known. In Section 9 we give the proofs for main results. In the Appendix we give all necessary technical results.

## 2 Main Conditions

In the paper we consider the estimation problem for the drift  $S$  on the interval  $[\mathbf{x}_0, \mathbf{x}_1]$ , where  $\mathbf{x}_0 < \mathbf{x}_1$  are some arbitrary fixed points. In order to obtain a reliable estimator of  $S$ , it is necessary to impose some conditions on this function which are similar to the periodicity of the deterministic signal in the white noise model. One of conditions which is sufficient for this purpose is the assumption that the process  $(y_t)_{t \geq 0}$  in (1.1) returns to any vicinity of each point  $x \in [\mathbf{x}_0, \mathbf{x}_1]$  infinite times. The ergodicity provides this property, when the coefficients of equation are known. In the case of unknown coefficients, one needs to impose a uniform ergodicity property. To obtain the

uniform ergodicity property for the process (1.1) we use the functional class introduced in [20], i.e. for any fixed  $\mathbf{L} \geq 1$ ,  $\mathbf{M} > 0$  and  $\mathbf{x}_* > |\mathbf{x}_0| + |\mathbf{x}_1|$  we set

$$\Sigma_{\mathbf{L},\mathbf{M}} = \left\{ S \in \mathbf{C}^1(\mathbb{R}) : \sup_{|x| \leq \mathbf{x}_*} (|S(x)| + |\dot{S}(x)|) \leq \mathbf{M}, \right. \\ \left. -\mathbf{L} \leq \inf_{|x| \geq \mathbf{x}_*} \dot{S}(x) \leq \sup_{|x| \geq \mathbf{x}_*} \dot{S}(x) \leq -1/\mathbf{L} \right\}. \quad (2.1)$$

Here and in the sequel we denote by  $\dot{f}$  and  $\ddot{f}$  the corresponding derivatives. Moreover, for some fixed parameters  $0 < \mathbf{b}_{min} \leq \mathbf{b}_{max} < \infty$  we denote by  $\mathcal{B}$  the class of functions  $b$  from  $\mathbf{C}^2(\mathbb{R})$  such that

$$\mathbf{b}_{min} \leq \inf_{x \in \mathbb{R}} |b(x)| \leq \sup_{x \in \mathbb{R}} \max (|b(x)|, |\dot{b}(x)|, |\ddot{b}(x)|) \leq \mathbf{b}_{max}. \quad (2.2)$$

Now we set

$$\Theta = \Sigma_{\mathbf{L},\mathbf{M}} \times \mathcal{B} = \{(S, b) : S \in \Sigma_{\mathbf{L},\mathbf{M}} \text{ and } b \in \mathcal{B}\}. \quad (2.3)$$

It is easy to see that the functions from  $\Sigma_{\mathbf{L},\mathbf{M}}$  are uniformly bounded on  $[\mathbf{x}_0, \mathbf{x}_1]$ , i.e.

$$s^* = \sup_{\mathbf{x}_0 \leq x \leq \mathbf{x}_1} \sup_{S \in \Sigma_{\mathbf{L},\mathbf{M}}} S^2(x) < \infty. \quad (2.4)$$

It should be noted that, for any  $\vartheta \in \Theta$ , (see, for example, [9]) the equation (1.1) has a unique strong solution and, moreover, there exists an invariant density for the process (1.1) which is defined as

$$\mathbf{q}_\vartheta(x) = \left( \int_{\mathbb{R}} b^{-2}(z) e^{\tilde{S}(z)} dz \right)^{-1} b^{-2}(x) e^{\tilde{S}(x)}, \quad (2.5)$$

where  $\tilde{S}(x) = 2 \int_0^x b^{-2}(v) S(v) dv$  (see, e.g., [22], Ch.4, 18, Th2). It is easy to see that this density is uniformly bounded on the class (2.3) and bounded away from zero on the interval  $[-\mathbf{x}_*, \mathbf{x}_*]$ :

$$0 < \mathbf{q}_* = \inf_{|x| \leq \mathbf{x}_*} \inf_{\vartheta \in \Theta} \mathbf{q}_\vartheta(x) \leq \sup_{x \in \mathbb{R}} \sup_{\vartheta \in \Theta} \mathbf{q}_\vartheta(x) = \mathbf{q}^* < +\infty \quad (2.6)$$

We need the following condition for the observation frequency.

**A<sub>1</sub>)** The frequency  $\delta$  in the observations (1.2) has the following form

$$\delta = \delta_T = \frac{1}{(T+1)l_T}, \quad (2.7)$$

where the function  $l_T$  is such that,

$$\lim_{T \rightarrow \infty} \frac{l_T}{\ln T} = +\infty. \quad (2.8)$$

For example, one can take  $l_T = (\ln T)^{1+\iota}$  for some  $\iota > 0$ .

**Remark 2.1.** *Note that we consider the efficient estimation problem only for the drift function  $S$ , i.e. the diffusion coefficient  $b$  is considered as a nuisance parameter.*

### 3 Point-wise estimators

In order to obtain a reliable estimator of the function  $S$  on the interval  $[\mathbf{x}_0, \mathbf{x}_1]$  we need some efficient point-wise estimators of this function. To give such estimators, we begin with the partition of the interval  $[\mathbf{x}_0, \mathbf{x}_1]$  by points  $(z_k)_{0 \leq k \leq n}$  defined as

$$z_k = \mathbf{x}_0 + \frac{k(\mathbf{x}_1 - \mathbf{x}_0)}{n}, \quad (3.1)$$

where  $n = n(T)$  is an integer-valued function of  $T$  such that

$$\lim_{T \rightarrow \infty} \frac{n(T)}{\sqrt{T}} = 1. \quad (3.2)$$

We can take, for example,

$$n = n(T) = 2 \left\lceil \frac{\sqrt{T}}{2} \right\rceil + 1, \quad (3.3)$$

where  $[x]$  is the integer part of the number  $x$ . To construct an efficient estimation procedure for  $S(z_k)$ , at any point  $z_k$ , we need to use some estimators of the invariant density  $\mathbf{q}_\theta(\cdot)$  and the squared diffusion coefficient  $b^2(\cdot)$ . To this end, we will estimate these both functions by making use of the first  $N_0$  observations, i.e.  $(y_{t_j})_{1 \leq j \leq N_0}$ . We set

$$N_0 = [N^\gamma(T)] \quad (3.4)$$

where  $5/6 < \gamma < 1$ .

### 3.1 Estimation of the invariant density

To estimate the density  $q_\vartheta$  we will make use of the following kernel estimator

$$\widehat{q}(z_k) = \frac{1}{2N_0 h_0} \sum_{j=1}^{N_0} \chi_{j,k}(h_0), \quad (3.5)$$

where  $h_0 = h_0(T) = T_0^{-1/2}$ ,  $T_0 = \delta N_0$ ,

$$\chi_{j,k}(h_0) = \chi\left(\frac{y_{t_{j-1}} - z_k}{h_0}\right) \quad \text{and} \quad \chi(y) = \mathbf{1}_{\{|y| \leq 1\}}.$$

Furthermore, we define a truncated estimator

$$\widetilde{q}(z_k) = \begin{cases} (v_T)^{1/2}, & \text{if } \widehat{q}(z_k) < (v_T)^{1/2}; \\ \widehat{q}(z_k), & \text{if } (v_T)^{1/2} \leq \widehat{q}(z_k) \leq (v_T)^{-1/2}; \\ (v_T)^{-1/2}, & \text{if } \widehat{q}(z_k) > (v_T)^{-1/2}, \end{cases} \quad (3.6)$$

where  $v_T$  is a strictly positive function of  $T$  verifying the following condition.

**A<sub>2</sub>)** Assume, that

$$\lim_{T \rightarrow \infty} \left( v_T + \frac{\ln T}{T(v_T)^2} + \frac{\ln T}{l_T(v_T)^5} \right) = 0. \quad (3.7)$$

For example, one can take  $v_T = \ln^{-\iota}(T+1)$  and  $l_T \geq \ln^{1+6\iota} T$ , for some  $\iota > 0$ .

**Proposition 3.1.** *Assume that the conditions **A<sub>1</sub>**)–**A<sub>2</sub>**) hold. Then, for any  $\mathbf{x}_0 < \mathbf{x}_1$  and for any  $a > 0$*

$$\lim_{T \rightarrow \infty} T^a \max_{\mathbf{x}_0 \leq x \leq \mathbf{x}_1} \sup_{\vartheta \in \Theta_{k,r}} \mathbf{P}_\vartheta(|\widetilde{q}_T(x) - q_\vartheta(x)| > v_T) = 0.$$

This proposition is shown in [20].

### 3.2 Estimation of the squared diffusion coefficient

To estimate the squared diffusion coefficient  $b^2(z_k)$  we use the following sequential procedure. First we define the sample size for this procedure :

$$\tau_{*,k} = \inf\left\{j \geq 1 : \sum_{l=1}^j \chi_{l,k}(h_0) \geq H_*\right\} \wedge N_0, \quad (3.8)$$



where  $H_* = h_0 N_0 / \ln(T + 1)$ . Then we set

$$\widehat{b}_k = \frac{\sum_{j=1}^{\tau_{*,k}} \chi_{j,k}(h_0) (y_{t_j} - y_{t_{j-1}})^2}{\delta H_*} \mathbf{1}_{\Gamma_{*,k}}, \quad (3.9)$$

where  $\Gamma_{*,k} = \{\sum_{j=1}^{N_0} \chi_{j,k}(h_0) \geq H_*\}$ .

This estimator satisfies the following property.

**Proposition 3.2.** *For any  $a > 0$ ,*

$$\limsup_{T \rightarrow \infty} \max_{\vartheta \in \Theta} T^{\gamma-1/2-a} \mathbf{E}_{\vartheta} |\widehat{b}_k - b^2(z_k)| = 0. \quad (3.10)$$

The proof is given in [21].

**Remark 3.1.** *Note that in the case of known diffusion coefficient  $b(\cdot)$ , we can take  $\widehat{b}_k = b^2(z_k)$ .*

### 3.3 Point-wise estimation of the drift coefficient

Now we estimate  $S(z_k)$ , at every point  $z_k$ , by making use of the observations  $(y_{t_j})_{N_0+1 \leq j \leq N}$ . To this end we use the sequential kernel estimator from [13] with the indicator kernel function  $\chi(y) = \mathbf{1}_{\{|y| \leq 1\}}$  and the duration of observations given by the stopping time

$$\tau_k = \inf\{l \geq N_0 + 1 : \sum_{j=N_0+1}^l \chi_{j,k}(h) \geq H_k\}, \quad (3.11)$$

where the bandwidth  $h = (\mathbf{x}_1 - \mathbf{x}_0)/2n$ ,

$$\chi_{j,k}(h) = \chi\left(\frac{y_{t_{j-1}} - z_k}{h}\right) \mathbf{1}_{\{1 \leq j \leq N\}} + \mathbf{1}_{\{j > N\}}$$

and  $H_k$  is some positive random threshold which will be specified later. It is clear that  $\tau_k < \infty$  a.s., for any  $H_k > 0$ . Now, we define the correction coefficient  $0 < \varkappa_k \leq 1$  as

$$\sum_{j=N_0+1}^{\tau_k-1} \chi_{j,k}(h) + \varkappa_k \chi_{\tau_k,k}(h) = H_k. \quad (3.12)$$

Moreover, using this coefficient we introduce the following weight sequence

$$\widetilde{\varkappa}_{j,k} = \mathbf{1}_{\{j < \tau_k\}} + \varkappa_k \mathbf{1}_{\{j = \tau_k\}}. \quad (3.13)$$

It should be noted, that for any  $j \geq 1$  the random variables  $\tilde{\varkappa}_{k,j}$  are  $\mathcal{G}_{j-1}$ -measurable, where  $\mathcal{G}_j = \sigma(y_{t_l}, 0 \leq l \leq j)$ . Now we define the sequential estimator for  $S(z_k)$  as

$$S_k^* = \frac{1}{\delta H_k} \sum_{j=N_0+1}^{\tau_k} \sqrt{\tilde{\varkappa}_{j,k}} \chi_{j,k}(h) \Delta y_{t_j} \mathbf{1}_{\Gamma_k}, \quad (3.14)$$

where  $\Delta y_{t_j} = y_{t_j} - y_{t_{j-1}}$  and

$$\Gamma_k = \{\tau_k \leq N\}. \quad (3.15)$$

By the same way used in the paper [20] to choose the threshold  $H_k$ , we use the truncated estimator  $\tilde{q}(z_k)$  given in (3.6) and we set

$$H_k = h(N - N_0)(2\tilde{q}(z_k) - v_T). \quad (3.16)$$

It should be noted that, as established in [10], this form for  $H_k$  provides the optimal convergence rate.

## 4 Regression model

In order to obtain an oracle inequality for discrete time data, we shall pass to a regression model by the same way as in [17].

Denote

$$\mathbf{G}_* = \cap_{k=1}^n \Gamma_k \quad \text{and} \quad Y_k = S_k^* \mathbf{1}_{\mathbf{G}_*}. \quad (4.1)$$

By making use of (1.1) and (3.14) we define a nonparametric regression model on the set  $\mathbf{G}_*$ :

$$Y_k = S(z_k) + \mathbf{g}_k + \sigma_k \xi_k, \quad \sigma_k = \frac{b(z_k)}{\sqrt{\delta H_k}}, \quad (4.2)$$

where

$$\begin{aligned} \mathbf{g}_k &= \frac{1}{\delta H_k} \sum_{j=N_0+1}^{\tau_k} \sqrt{\tilde{\varkappa}_{j,k}} \chi_{j,k}(h) \int_{t_{j-1}}^{t_j} S(y_u) du - S(z_k) \\ &+ \frac{1}{\delta H_k} \sum_{j=N_0+1}^{\tau_k} \sqrt{\tilde{\varkappa}_{j,k}} \chi_{j,k}(h) \int_{t_{j-1}}^{t_j} (b(y_s) - b(z_k)) dW_s \end{aligned} \quad (4.3)$$

and

$$\xi_k = \frac{1}{\sqrt{\delta H_k}} \sum_{j=N_0+1}^{\tau_k} \sqrt{\tilde{\chi}_{j,k}} \chi_{j,k}(h) \Delta W_{t_j}. \quad (4.4)$$

Note that the coefficients  $(\sigma_l)_{1 \leq l \leq n}$  are random variables and using their definitions one can obtain the following bounds

$$\sigma_{0,*} \leq \min_{1 \leq l \leq n} \sigma_l^2 \leq \max_{1 \leq l \leq n} \sigma_l^2 \leq \sigma_{1,*}, \quad (4.5)$$

where

$$\sigma_{0,*} = \frac{v_T \mathbf{b}_{min}}{\delta N h} \quad \text{and} \quad \sigma_{1,*} = \frac{\mathbf{b}_{max}}{v_T \delta (N - N_0) h}.$$

We estimate the parameter  $\sigma_l^2$  as follows:

$$\hat{\sigma}_l = \frac{\hat{b}_l}{\delta H_l}, \quad (4.6)$$

where  $\hat{b}_l$  is the estimator of the squared diffusion coefficient  $b^2(z_l)$  defined in (3.9). In order to obtain the oracle inequality, we need to study the properties of the last estimator. To this end we set

$$\varpi_T^* = n \max_{1 \leq l \leq n} \mathbf{E}_\vartheta |\hat{\sigma}_l - \sigma_l^2|. \quad (4.7)$$

**Proposition 4.1.** *Assume that the conditions  $\mathbf{A}_1)$  –  $\mathbf{A}_2)$  hold. Then, for any  $a > 0$ ,*

$$\lim_{T \rightarrow \infty} T^{\gamma-1/2-a} \varpi_T^* = 0. \quad (4.8)$$

This proposition is proved in [21].

## 5 Model selection

First we choose a basis  $(\phi_j)_{j \geq 1}$  in  $\mathcal{L}_2([\mathbf{x}_0, \mathbf{x}_1])$  such that, for any  $1 \leq i, j \leq n$ ,

$$(\phi_i, \phi_j)_n = \frac{\mathbf{x}_1 - \mathbf{x}_0}{n} \sum_{l=1}^n \phi_i(z_l) \phi_j(z_l) = \mathbf{1}_{\{i=j\}}. \quad (5.1)$$

For example, one can take the trigonometric basis defined as  $\phi_1(x) \equiv 1/\sqrt{\mathbf{x}_1 - \mathbf{x}_0}$  and, for  $j \geq 2$ ,

$$\phi_j(x) = \sqrt{\frac{2}{\mathbf{x}_1 - \mathbf{x}_0}} \begin{cases} \cos(2\pi[j/2]l_0(x)) & \text{for even } j; \\ \sin(2\pi[j/2]l_0(x)) & \text{for odd } j, \end{cases} \quad (5.2)$$

where  $[a]$  denotes the integer part of  $a$  and  $l_0(x) = (x - \mathbf{x}_0)/(\mathbf{x}_1 - \mathbf{x}_0)$ . Note that if  $n$  is odd, then this basis is orthonormal for the empirical inner product, i.e. satisfies the property (5.1). In the sequel, we assume that the  $n$  is odd, i.e. as in (3.3). By making use of this property we define the discrete Fourier representation for  $S$  on the sieve (3.1), i.e.,

$$S(z_k) = \sum_{j=1}^n \theta_{j,n} \phi_j(z_k), \quad 1 \leq k \leq n, \quad (5.3)$$

where

$$\theta_{j,n} = (S, \phi_j)_n = \frac{\mathbf{x}_1 - \mathbf{x}_0}{n} \sum_{l=1}^n S(z_l) \phi_j(z_l).$$

Moreover, using the regression model (4.2) we estimate these coefficients as

$$\hat{\theta}_{j,n} = (Y, \phi_j)_n = \frac{\mathbf{x}_1 - \mathbf{x}_0}{n} \sum_{l=1}^n Y_l \phi_j(z_l). \quad (5.4)$$

By the model (4.2), we obtain on the set  $\mathbf{G}_*$

$$\hat{\theta}_{j,n} = \theta_{j,n} + \zeta_{j,n}, \quad \zeta_{j,n} = \mathbf{g}_{j,n} + \sqrt{\frac{\mathbf{x}_1 - \mathbf{x}_0}{n}} \xi_{j,n}, \quad (5.5)$$

where

$$\xi_{j,n} = \sqrt{\frac{\mathbf{x}_1 - \mathbf{x}_0}{n}} \sum_{l=1}^n \sigma_l \xi_l \phi_j(z_l), \quad \mathbf{g}_{j,n} = \frac{\mathbf{x}_1 - \mathbf{x}_0}{n} \sum_{l=1}^n \mathbf{g}_l \phi_j(z_l).$$

We estimate the values  $S(z_k), 1 \leq k \leq n$ , by the weighted least squares estimators

$$\hat{S}_\lambda(z_k) = \sum_{j=1}^n \lambda(j) \hat{\theta}_{j,n} \phi_j(z_k), \quad 1 \leq k \leq n, \quad (5.6)$$

where the weight vector  $\lambda = (\lambda(1), \dots, \lambda(n))'$  belongs to some finite set  $\Lambda$  from  $[0, 1]^n$ . In the sequel, we denote by  $\nu$  the cardinal number of the set  $\Lambda$ ,  $\nu = \text{card}(\Lambda)$ , which is a function of  $T$ , i.e.  $\nu = \nu_T$ . Moreover, we set the norm for  $\Lambda$  as

$$\Lambda_* = \max_{\lambda \in \Lambda} \sum_{j=1}^n \lambda(j) \quad (5.7)$$

which can be a function of  $T$ , i.e.  $\Lambda_* = \Lambda_*(T)$ . We assume that the basis functions and the weight set  $\Lambda$  satisfy the following condition.

**A<sub>3</sub>)** For any  $a > 0$ ,

$$\lim_{T \rightarrow \infty} \frac{\nu_T}{T^a} = 0 \quad \text{and} \quad \lim_{T \rightarrow \infty} \frac{\Lambda_*(T)}{T^{1/3+a}} = 0. \quad (5.8)$$

To estimate the function  $S$  on the interval  $[\mathbf{x}_0, \mathbf{x}_1]$ , we use the step-function approximation, i.e.,

$$\widehat{S}_\lambda(x) = \sum_{l=1}^n \widehat{S}_\lambda(z_l) \mathbf{1}_{\{z_{l-1} < x \leq z_l\}}, \quad x \in [\mathbf{x}_0, \mathbf{x}_1]. \quad (5.9)$$

Now one needs to choose a cost function in order to define an optimal weight  $\lambda \in \Lambda$ . A best candidate for the cost function should be the empirical squared error given by the relation

$$\text{Err}_n(\lambda) = \|\widehat{S}_\lambda - S\|_n^2 \rightarrow \min.$$

In our case, the empirical squared error is equal to

$$\text{Err}_n(\lambda) = \sum_{j=1}^n \lambda^2(j) \widehat{\theta}_{j,n}^2 - 2 \sum_{j=1}^n \lambda(j) \widehat{\theta}_{j,n} \theta_{j,n} + \sum_{j=1}^n \theta_{j,n}^2. \quad (5.10)$$

Since coefficients  $\theta_{j,n}$  are unknown, we need to replace the term  $\widehat{\theta}_{j,n} \theta_{j,n}$  by some estimator which we choose as

$$\widetilde{\theta}_{j,n} = \widehat{\theta}_{j,n}^2 - \frac{\mathbf{x}_1 - \mathbf{x}_0}{n} \widehat{\sigma}_{j,n} \quad \text{and} \quad \widehat{\sigma}_{j,n} = \frac{\mathbf{x}_1 - \mathbf{x}_0}{n} \sum_{l=1}^n \widehat{\sigma}_l \phi_j^2(z_l), \quad (5.11)$$

where  $\widehat{\sigma}_l$  is the estimator for  $\sigma_l^2$  defined in (4.6). Note that if the diffusion is known, then we take in (5.11)  $\widehat{\sigma}_{j,n} = \sigma_{j,n}$  and

$$\sigma_{j,n} = \frac{\mathbf{x}_1 - \mathbf{x}_0}{n} \sum_{l=1}^n \sigma_l^2 \phi_j^2(z_l). \quad (5.12)$$

It is clear that the inequalities (4.5) imply

$$\sigma_{0,*} \leq \min_{1 \leq l \leq n} \sigma_{l,n} \leq \max_{1 \leq l \leq n} \sigma_{l,n} \leq \sigma_{1,*}. \quad (5.13)$$

Now, for using the estimator (5.11) instead of  $\theta_{j,n} \widehat{\theta}_{j,n}$  one needs to add to the cost function a suitable penalty term that we take as

$$\widehat{P}_n(\lambda) = \frac{\mathbf{x}_1 - \mathbf{x}_0}{n} \sum_{j=1}^n \lambda^2(j) \widehat{\sigma}_{j,n} \quad (5.14)$$

if the diffusion is unknown and as

$$P_n(\lambda) = \frac{\mathbf{x}_1 - \mathbf{x}_0}{n} \sum_{j=1}^n \lambda^2(j) \sigma_{j,n} \quad (5.15)$$

when the diffusion is known. Finally, we use the following cost function

$$J_n(\lambda) = \sum_{j=1}^n \lambda^2(j) \widehat{\theta}_{j,n}^2 - 2 \sum_{j=1}^n \lambda(j) \widetilde{\theta}_{j,n} + \rho \widehat{P}_n(\lambda), \quad (5.16)$$

where the positive coefficient  $0 < \rho < 1$  will be specified later.

We define the model selection procedure as

$$\widehat{\lambda} = \operatorname{argmin}_{\lambda \in \Lambda} J_n(\lambda) \quad \text{and} \quad \widehat{S}_* = \widehat{S}_{\widehat{\lambda}}. \quad (5.17)$$

**Remark 5.1.** *It should be emphasized that if in the model (1.1) the diffusion coefficient  $b(\cdot)$  is known, then the model selection procedure (5.17) is defined through minimizing the cost function  $J_n(\lambda)$  with the penalty term (5.15).*

For the risk (1.3), we have shown the following result in [21].

**Theorem 5.1.** *Assume that the conditions  $\mathbf{A}_1) - \mathbf{A}_3)$  hold. Then, for any  $T \geq 1$ ,  $0 < \rho \leq 1/8$  and  $\vartheta \in \Theta$ , the estimation procedure  $\widehat{S}_*$  defined in (5.17) satisfies the inequality*

$$\mathcal{R}_\vartheta(\widehat{S}_*) \leq \frac{(1 + \rho)^2(1 + 5\rho)}{1 - 6\rho} \min_{\lambda \in \Lambda} \mathcal{R}_\vartheta(\widehat{S}_\lambda) + \frac{\mathbf{U}_{\vartheta,T}}{\rho T}, \quad (5.18)$$

where the remainder term  $\mathbf{U}_{\vartheta,T}$  is such that for any  $a > 0$ ,

$$\lim_{T \rightarrow \infty} T^{-a} \sup_{\vartheta \in \Theta} \mathbf{U}_{\vartheta,T} = 0. \quad (5.19)$$

In the sequel to obtain the efficient properties for this procedure we will use the special weight coefficients introduced in [15, 16]. To this end we consider a 2-dimensional numerical grid of the form

$$\mathcal{A} = \{1, \dots, \mathbf{k}^*\} \times \{r_1, \dots, r_{m^*}\}, \quad (5.20)$$

where  $r_i = i\varepsilon$  and  $m^* = \lceil 1/\varepsilon^2 \rceil$ . The both parameters  $\mathbf{k}^* \geq 1$  and  $0 < \varepsilon \leq 1$  are some functions of  $T$ , i.e.  $\mathbf{k}^* = \mathbf{k}_T^*$  and  $\varepsilon = \varepsilon_T$ , such that, for any  $\gamma > 0$ ,

$$\lim_{T \rightarrow \infty} \left( \varepsilon_T + \frac{1}{T^\gamma \varepsilon_T} + \frac{1}{\mathbf{k}_T^*} + \frac{\mathbf{k}_T^*}{\ln T} \right) = 0. \quad (5.21)$$

One can take, for example,

$$\varepsilon_T = \frac{1}{\ln(T+1)} \quad \text{and} \quad \mathbf{k}^* = \bar{k} + \sqrt{\ln(T+1)},$$

for some fixed  $\bar{k} \geq 1$ . For any  $\alpha = (k, l) \in \mathcal{A}$ , we introduce the weight sequence  $\lambda_\alpha = (\lambda_\alpha(j))_{j \geq 1}$  as

$$\lambda_\alpha(j) = \mathbf{1}_{\{1 \leq j \leq j_0\}} + (1 - (j/\omega_\alpha)^k) \mathbf{1}_{\{j_0 < j \leq \omega_\alpha\}}, \quad (5.22)$$

where  $j_0 = j_0(\alpha) = \lfloor \omega_\alpha / \ln(T+1) \rfloor$ ,  $\omega_\alpha = \check{\omega}_k (lT)^{1/(2k+1)}$  and

$$\check{\omega}_k = (\mathbf{x}_1 - \mathbf{x}_0) \left( \frac{(k+1)(2k+1)}{\pi^{2k} k} \right)^{1/(2k+1)}.$$

We set

$$\Lambda = (\lambda_\alpha)_{\alpha \in \mathcal{A}}. \quad (5.23)$$

Note that, in this case, the cardinal  $\nu$  of the set  $\Lambda$  is the function of  $T$ , i.e.  $\nu = \nu_T = \mathbf{k}^* m^*$  and the conditions (5.21) imply that, for any  $a > 0$ ,

$$\lim_{T \rightarrow \infty} \frac{\nu_T}{T^a} = 0. \quad (5.24)$$

Moreover, from the definition (5.22) we can obtain that, for any  $\alpha \in \mathcal{A}$ ,

$$\sum_{j=1}^n \lambda_\alpha(j) \leq \omega_\alpha \leq \check{\omega}_k \left( \frac{T}{\varepsilon_T} \right)^{1/3}.$$

Therefore, for any  $a > 0$ ,

$$\lim_{T \rightarrow \infty} \frac{\Lambda^*}{T^{1/3+a}} = 0. \quad (5.25)$$

Hence, the condition  $\mathbf{A}_3$ ) holds and we obtain the following result.

**Theorem 5.2.** *Assume that the conditions  $\mathbf{A}_1$ ) –  $\mathbf{A}_2$ ) hold. Then, the model selection procedure (5.17) with the weights (5.23) satisfies the oracle inequality (5.18) with the remainder term satisfying the property (5.19), for any  $a > 0$ .*

**Remark 5.2.** *It should be noted that similarly to [17], we will use the inequality (5.18) to provide the efficiency property in the adaptive setting. This means that without using the regularity properties of the unknown function  $S$  we can estimate from above the risk for the model selection procedure by the risk for the efficient estimation procedure constructed on the regularity properties of the function  $S$ . The upper bound for the risk of the model selection procedure follows from the sharp oracle inequality.*

## 6 Main results

In order to study the asymptotic efficiency of estimators we use a suitable class of drift coefficients  $S$ . To this end we define the following functional Sobolev ball

$$W_{k,r}^0 = \{f \in \mathbf{C}_0^k([\mathbf{x}_0, \mathbf{x}_1]) : \sum_{j=0}^k \|f^{(j)}\|^2 \leq r\}, \quad (6.1)$$

where  $r > 0$  and the integer  $k \geq 1$  are some parameters,  $\mathbf{C}_0^k([\mathbf{x}_0, \mathbf{x}_1])$  is the space of  $k$  times differentiable functions  $f : \mathbb{R} \rightarrow \mathbb{R}$  such that

$$f^{(i)}(x) = 0, \quad \text{for } 0 \leq i \leq k-1 \quad \text{and} \quad x \notin [\mathbf{x}_0, \mathbf{x}_1].$$

Moreover, let  $S_0$  be a fixed continuously differentiable function from  $\Sigma_{\mathbf{L},\mathbf{M}}$  such that  $S_0(x) = 0$  for  $\mathbf{x}_0 \leq x \leq \mathbf{x}_1$ . We set

$$W_{k,r} = S_0 + W_{k,r}^0 \quad \text{and} \quad \Theta_{k,r} = W_{k,r} \times \mathcal{B}. \quad (6.2)$$

Note that one can represent the functional class  $W_{k,r}^0$  as an ellipse in  $\mathcal{L}_2([\mathbf{x}_0, \mathbf{x}_1])$  with trigonometric basis  $(\phi_j)_{j \geq 0}$ :

$$W_{k,r}^0 = \{f \in \mathbf{C}_0^k([\mathbf{x}_0, \mathbf{x}_1]) : \sum_{j=0}^{\infty} a_j \theta_j^2 \leq r\}, \quad (6.3)$$

where  $\theta_j = (f, \phi_j) = \int_{\mathbf{x}_0}^{\mathbf{x}_1} f(x) \phi_j(x) dx$  and

$$a_j = \sum_{i=0}^k \left( \frac{2\pi[j/2]}{\mathbf{x}_1 - \mathbf{x}_0} \right)^{2i}.$$

In order to formulate our asymptotic results, we define, for any  $\vartheta \in \Theta_{k,r}$ , the following normalizing coefficient  $\gamma(\vartheta)$

$$\gamma(\vartheta) = l_* \mathbf{J}_{\vartheta}^{k_*}, \quad k_* = \frac{2k}{2k+1}, \quad (6.4)$$

where

$$l_* = (2k+1)r^{1-k_*} \left( \frac{k}{\pi(k+1)(2k+1)} \right)^{k_*}$$

and

$$\mathbf{J}_{\vartheta} = \int_{\mathbf{x}_0}^{\mathbf{x}_1} \frac{b^2(x)}{\mathbf{q}_{\vartheta}(x)} dx. \quad (6.5)$$



It is well known that, for any  $S \in \Theta_{k,r}$ , the optimal rate of convergence of estimators is  $T^{-2k/(2k+1)}$  (see, for example, [11]).

Now we denote by  $\Xi_T$  the set of all possible estimators of  $S$  which are measurable with respect to the  $\sigma$ -field  $\sigma\{y_t, 0 \leq t \leq T\}$ , i.e. based on the observations  $(y_t)_{0 \leq t \leq T}$ .

**Theorem 6.1.** *For any integer  $k \geq 1$  and  $r > 0$ , the quadratic risk  $\mathcal{R}_\vartheta(\widehat{S})$  admits the following asymptotic lower bound*

$$\liminf_{T \rightarrow \infty} T^{k_*} \inf_{\widehat{S} \in \Xi_T} \sup_{\vartheta \in \Theta_{k,r}} v(\vartheta) \mathcal{R}_\vartheta(\widehat{S}) \geq l_*,$$

where  $v(\vartheta) = \mathbf{J}_\vartheta^{-k_*}$ .

Assume now, that the penalty parameter  $\rho$  in (5.16) is a function of  $T$ , i.e.  $\rho = \rho_T$  such that, for any  $a > 0$ ,

$$\lim_{T \rightarrow \infty} \rho_T = 0 \quad \text{and} \quad \lim_{T \rightarrow \infty} T^a \rho_T = \infty. \quad (6.6)$$

We can take, for example,  $\rho_T = (6 + \ln(T + 1))^{-1}$ .

**Theorem 6.2.** *Assume that the conditions  $\mathbf{A}_1) - \mathbf{A}_2)$  hold. Then, for any integer  $k \geq 2$  and  $r > 0$ , the quadratic risk of the estimation procedure  $\widehat{S}_*$  defined through the trigonometric basis (5.2) in (5.17) and (5.23) with the parameter  $\rho$  of the form (6.6) satisfies the following upper bound*

$$\limsup_{T \rightarrow \infty} T^{k_*} \sup_{\vartheta \in \Theta_{k,r}} v(\vartheta) \mathcal{R}_\vartheta(\widehat{S}_*) \leq l_*.$$

Theorems 6.1 – 6.2 imply immediately the efficiency property :

**Theorem 6.3.** *Assume that the conditions  $\mathbf{A}_1) - \mathbf{A}_2)$  hold. Then the estimation procedure  $\widehat{S}_*$  defined through the trigonometric basis (5.2) in (5.17) and (5.23) with the parameter  $\rho$  of the form (6.6) is asymptotically efficient, i.e.*

$$\lim_{T \rightarrow \infty} \frac{\inf_{\widehat{S} \in \Xi_T} \sup_{\vartheta \in \Theta_{k,r}} v(\vartheta) \mathcal{R}_\vartheta(\widehat{S})}{\sup_{\vartheta \in \Theta_{k,r}} v(\vartheta) \mathcal{R}_\vartheta(\widehat{S}_*)} = 1.$$

**Remark 6.1.** *It should be noted that from Theorems 6.1 – 6.2 it follows that the function (6.4) is the Pinsker constant, which is calculated for the first time for the model (1.1). For the model (1.1) with the diffusion coefficient  $b(\cdot) \equiv 1$  the Pinsker constant was calculated in [17] which in this case coincides with (6.4). Note also that the parameter  $l_*$  is the well - known Pinsker constant for the "signal plus white noise" model obtained in [33]. Therefore, the Pinsker constant for the model (1.1) is obtained by multiplying the constant from [33] by the Pinsker variance (6.5) in the power  $k_*$ .*

## 7 Properties of the regression model (4.2)

First we recall the main properties of the model (4.2) from [21].

The random variables  $(\xi_k)_{1 \leq k \leq n}$  from (5.5), satisfy the following property.

**Proposition 7.1.** *The random variables  $(\xi_k)_{1 \leq k \leq n}$  are  $\mathcal{N}(0, 1)$  i.i.d. conditionally to the  $\sigma$ -field  $\mathcal{G}_{N_0}$ .*

**Proposition 7.2.** *Assume that the conditions  $\mathbf{A}_1)$  –  $\mathbf{A}_2)$  hold. Then, for any  $a > 0$ ,*

$$\lim_{T \rightarrow \infty} T^a \sup_{\vartheta \in \Theta} \mathbf{P}_\vartheta(\mathbf{G}_*^c) = 0. \quad (7.1)$$

These propositions are proved in [21].

Now, we study the heteroscedastic property in the model (4.2). To this end we set

$$\mathbf{s}_n = (\mathbf{x}_1 - \mathbf{x}_0) \sum_{l=1}^n \sigma_l^2, \quad (7.2)$$

where the variances  $(\sigma_l)_{1 \leq l \leq n}$  are defined in (4.2).

**Proposition 7.3.** *Assume that the condition  $\mathbf{A}_1)$  holds. Then*

$$\lim_{T \rightarrow \infty} \sup_{\vartheta \in \Theta_{k,r}} \mathbf{E}_\vartheta \left| \mathbf{s}_n - \frac{\mathbf{J}_\vartheta}{\mathbf{x}_1 - \mathbf{x}_0} \right| = 0. \quad (7.3)$$

**Proof.** Using the definition of  $\sigma_l$  in (4.2) and taking account of the form of  $h$  given in (3.11), we can represent the quantity  $\mathbf{s}_n$  as

$$\begin{aligned} \mathbf{s}_n &= \sum_{l=1}^n \frac{nb^2(z_l)}{\delta h(N - N_0)(2\tilde{q}_T(z_l) - v_T)} (z_l - z_{l-1}) \\ &= \frac{1}{\mathbf{x}_1 - \mathbf{x}_0} \sum_{l=1}^n \tilde{b}_\vartheta(z_l)(z_l - z_{l-1}) + R_1(\vartheta) + R_2(\vartheta), \end{aligned} \quad (7.4)$$

where  $\tilde{b}_\vartheta(x) = b^2(x)/\mathbf{q}_\vartheta(x)$ ,

$$R_1(\vartheta) = \frac{1}{\mathbf{x}_1 - \mathbf{x}_0} \left( \frac{n^2}{\delta(N - N_0)} - 1 \right) \sum_{l=1}^n \tilde{b}_\vartheta(z_l)(z_l - z_{l-1})$$

and

$$R_2(\vartheta) = \sum_{l=1}^n \frac{nb^2(x_l)}{\delta h(N - N_0)} \left( \frac{1}{2\tilde{q}_T(z_l) - v_T} - \frac{1}{2\mathbf{q}_\vartheta(z_l)} \right) (z_l - z_{l-1}).$$

First, note, that the function  $\tilde{b}_\vartheta(\cdot)$  and its derivative are uniformly bounded, i.e.

$$\sup_{\vartheta \in \Theta_{k,r}} \max_{\mathbf{x}_0 \leq z \leq \mathbf{x}_1} \left( \tilde{b}_\vartheta(z) + |\tilde{b}'_\vartheta(z)| \right) < \infty. \quad (7.5)$$

Therefore,

$$\lim_{T \rightarrow \infty} \sup_{\vartheta \in \Theta_{k,r}} \left| \sum_{l=1}^n \tilde{b}_\vartheta(z_l)(z_l - z_{l-1}) - \mathbf{J}_\vartheta \right| = 0.$$

As the second term (7.4), note that, in view of the condition (3.2),

$$\lim_{T \rightarrow \infty} \frac{n^2}{\delta(N - N_0)} = 1.$$

Therefore, the property (7.5) provides that

$$\lim_{T \rightarrow \infty} \sup_{\vartheta \in \Theta_{k,r}} |R_1(\vartheta)| = 0.$$

Moreover, taking into account that  $2\tilde{q}_T(z_l) - v_T > v_T^{1/2}$ , we obtain, for sufficiently large  $T$ , that for some  $C > 0$

$$|R_2(\vartheta)| \leq C \left( \sum_{l=1}^n v_T^{-1/2} (|\tilde{q}_T(z_l) - \mathbf{q}_\vartheta(z_l)|) (z_l - z_{l-1}) + \sqrt{v_T} \right).$$

Note here, that for any  $1 \leq l \leq n$  and for sufficiently large  $T$ ,

$$v_T^{-1/2} \mathbf{E}_\vartheta |\tilde{q}_T(z_l) - \mathbf{q}_\vartheta(z_l)| \leq 2v_T^{-1} \mathbf{P}_\vartheta (|\tilde{q}_T(z_l) - \mathbf{q}_\vartheta(z_l)| > v_T) + \sqrt{v_T}.$$

Therefore, due to Proposition 3.1,

$$\lim_{T \rightarrow \infty} \sup_{\vartheta \in \Theta_{k,r}} |R_2(\vartheta)| = 0.$$

Hence Proposition 7.3.  $\square$

## 8 Non adaptive estimation

We start with the non adaptive case, i.e. we assume that the parameters  $k$ ,  $r$  and the function  $\mathbf{J}_\vartheta$  are known. In this case, we use the weight estimator (5.9) from the family (5.23)

$$\tilde{S} = \tilde{S}_{\tilde{\lambda}} \quad \text{and} \quad \tilde{\lambda} = \lambda_{\tilde{\alpha}}, \quad (8.1)$$

where  $\tilde{\alpha} = (k, \tilde{l})$ ,  $\tilde{l} = \tilde{l}_T = [\bar{r}(\vartheta)/\varepsilon] \varepsilon$  and  $\bar{r}(\vartheta) = r/\mathbf{J}_\vartheta$ . We remind that  $\varepsilon = \varepsilon_T = 1/\ln(T+1)$ . Firstly, we need to study the asymptotic properties of the weights  $\tilde{\lambda}$ . To this end we set

$$\tilde{\Upsilon}_T(\vartheta) = \sum_{j=1}^n (1 - \tilde{\lambda}(j))^2 \theta_{j,n}^2 + \frac{\mathbf{J}_\vartheta}{(\mathbf{x}_1 - \mathbf{x}_0)n^2} \sum_{j=1}^n \tilde{\lambda}^2(j). \quad (8.2)$$

**Proposition 8.1.** *For any  $k \geq 2$  and  $r > 0$ ,*

$$\limsup_{T \rightarrow \infty} T^{k_*} \sup_{\vartheta \in \Theta_{k,r}} v(\vartheta) \tilde{\Upsilon}_T(\vartheta) \leq l_*. \quad (8.3)$$

**Proof.** First of all, note that

$$0 < \inf_{\vartheta \in \Theta_{k,r}} \mathbf{J}_\vartheta \leq \sup_{\vartheta \in \Theta_{k,r}} \mathbf{J}_\vartheta < \infty. \quad (8.4)$$

This implies directly that

$$\lim_{T \rightarrow \infty} \sup_{\vartheta \in \Theta_{k,r}} \left| \frac{\tilde{l}_T}{\bar{r}(\vartheta)} - 1 \right| = 0, \quad (8.5)$$

where  $\bar{r}(\vartheta) = r/\mathbf{J}_\vartheta$ . Moreover, note that

$$T^{k_*} v(\vartheta) \tilde{\Upsilon}_T(\vartheta) \leq T^{k_*} v(\vartheta) \mathbf{S}_T + \frac{(\mathbf{J}_\vartheta)^{1-k_*}}{T^{1-k_*}(\mathbf{x}_1 - \mathbf{x}_0)} \sum_{j=1}^n \tilde{\lambda}^2(j), \quad (8.6)$$

where  $\mathbf{S}_T = \sum_{j=1}^n (1 - \tilde{\lambda}(j))^2 \theta_{j,n}^2$ . We decompose  $\mathbf{S}_T$  as

$$\mathbf{S}_T = \mathbf{S}_{1,T} + \mathbf{S}_{2,T}$$

with

$$\mathbf{S}_{1,T} = \sum_{j=j_0+1}^{[\tilde{\omega}]} (1 - \tilde{\lambda}(j))^2 \theta_{j,n}^2 \quad \text{and} \quad \mathbf{S}_{2,T} = \sum_{j=[\tilde{\omega}]+1}^n \theta_{j,n}^2.$$

Let us recall that

$$\tilde{\omega} = \omega_{\tilde{\alpha}} = \check{\omega}_k \left( T \tilde{l}_T \right)^{1/(2k+1)}.$$

Lemma A.2 yields

$$\mathbf{S}_{1,T} \leq (1 + \tilde{\varepsilon}) \sum_{j=j_0}^{[\tilde{\omega}]} (1 - \tilde{\lambda}(j))^2 \theta_j^2 + 2r(1 + \tilde{\varepsilon}^{-1}) \frac{\tilde{\omega}}{n^{2k}},$$

and Lemma A.3 yields

$$\mathbf{S}_{2,T} \leq (1 + \tilde{\varepsilon}) \sum_{j > \tilde{\omega}} \theta_j^2 + (1 + \tilde{\varepsilon}^{-1}) \frac{r}{n^2 \tilde{\omega}^{2(k-1)}}.$$

These inequalities imply that

$$\mathbf{S}_T \leq (1 + \tilde{\varepsilon}) \mathbf{S}_T^* + 2r(1 + \tilde{\varepsilon}^{-1}) \tilde{\omega}_T, \quad (8.7)$$

where

$$\mathbf{S}_T^* = \sum_{j \geq 1} (1 - \tilde{\lambda}(j))^2 \theta_j^2 = \sum_{j \leq \tilde{\omega}} (1 - \tilde{\lambda}(j))^2 \theta_j^2 + \sum_{j > \tilde{\omega}} \theta_j^2 := \mathbf{S}_{1,T}^* + \mathbf{S}_{2,T}^*$$

and

$$\tilde{\omega}_T = \frac{\tilde{\omega}}{n^{2k}} + \frac{1}{n^2 \tilde{\omega}^{2(k-1)}}.$$

Note, that

$$\begin{aligned} T^{k_*} v(\vartheta) \mathbf{S}_{1,T}^* &= T^{k_*} v(\vartheta) \sum_{j=j_0}^{[\tilde{\omega}]} (1 - \tilde{\lambda}(j))^2 \theta_j^2 = \frac{v(\vartheta)}{\tilde{\omega}_k^{2k} (\tilde{l}_T)^{k_*}} \sum_{j=j_0}^{[\tilde{\omega}]} j^{2k} \theta_j^2 \\ &\leq \frac{v(\vartheta)}{\tilde{\omega}_k^{2k} (\tilde{l}_T)^{k_*}} \varpi_{j_0} \sum_{j=j_0}^{[\tilde{\omega}]} a_j \theta_j^2, \end{aligned}$$

where  $\varpi_n = \sup_{j \geq n} j^{2k}/a_j$ . It is clear that

$$\lim_{n \rightarrow \infty} \varpi_n = \frac{(\mathbf{x}_1 - \mathbf{x}_0)^{2k}}{\pi^{2k}}.$$

Therefore, from (8.5) we obtain that

$$\limsup_{n \rightarrow \infty} \sup_{\vartheta \in \Theta_{k,r}} \frac{T^{k_*} v(\vartheta) \mathbf{S}_{1,T}^*}{\sum_{j=j_0}^{[\tilde{\omega}]} a_j \theta_j^2} \leq \frac{(\mathbf{x}_1 - \mathbf{x}_0)^{2k}}{\pi^{2k} \tilde{\omega}_k^{2k} r^{k_*}} = \left( \frac{k}{(2k+1)(k+1)r\pi} \right)^{k_*}.$$

Futher note, that for any  $0 < \tilde{\varepsilon} < 1$  and for sufficiently large  $T$ ,

$$\begin{aligned} \mathbf{S}_{2,T}^* &= \sum_{j > \tilde{\omega}} \theta_j^2 \leq \frac{(\mathbf{x}_1 - \mathbf{x}_0)^{2k}}{\pi^{2k} \tilde{\omega}^{2k}} \sum_{j > \tilde{\omega}} a_j \theta_j^2 = \frac{(\mathbf{x}_1 - \mathbf{x}_0)^{2k}}{\pi^{2k} (T \tilde{l}_T)^{k_*} \tilde{\omega}_k^{2k}} \sum_{j > \tilde{\omega}} a_j \theta_j^2 \\ &\leq (1 + \tilde{\varepsilon}) \frac{J_{\vartheta}^{k_*} (\mathbf{x}_1 - \mathbf{x}_0)^{2k}}{\pi^{2k} (rT)^{k_*} \tilde{\omega}_k^{2k}} \sum_{j > \tilde{\omega}} a_j \theta_j^2. \end{aligned}$$

This inequality implies

$$\limsup_{n \rightarrow \infty} \sup_{\vartheta \in \Theta_{k,r}} \frac{T^{k_*} v(\vartheta) \mathbf{S}_{2,T}^*}{\sum_{j > \tilde{\omega}} a_j \theta_j^2} \leq \frac{(\mathbf{x}_1 - \mathbf{x}_0)^{2k}}{\pi^{2k} \tilde{\omega}_k^{2k} r^{k_*}} = \left( \frac{k}{(2k+1)(k+1)r\pi} \right)^{k_*}.$$

So, taking into account in (8.7) that for  $k \geq 2$

$$\lim_{T \rightarrow \infty} \sup_{S \in W_{k,r}} T^{k_*} \tilde{\omega}_T = 0,$$

we obtain that

$$\limsup_{T \rightarrow \infty} \sup_{\vartheta \in \Theta_{k,r}} T^{k_*} v(\vartheta) \mathbf{S}_T \leq r^{1-k_*} \left( \frac{k}{(2k+1)(k+1)\pi} \right)^{k_*}. \quad (8.8)$$

Moreover, we can check directly that

$$\limsup_{T \rightarrow \infty} \sup_{\vartheta \in \Theta_{k,r}} \frac{\mathbf{J}_\vartheta^{1-k_*}}{T^{1-k_*} (\mathbf{x}_1 - \mathbf{x}_0)} \sum_{j=1}^n \tilde{\lambda}^2(j) \leq \mathbf{u}_k^*, \quad (8.9)$$

where

$$\mathbf{u}_k^* = \frac{\tilde{\omega}_k r^{1-k_*} 2k^2}{(\mathbf{x}_1 - \mathbf{x}_0)(2k+1)(k+1)} = 2kr^{1-k_*} \left( \frac{k}{(2k+1)(k+1)\pi} \right)^{k_*}.$$

Using now (8.8) and (8.9) in (8.6) we obtain the limit equality (8.3) and, hence Proposition 8.1.  $\square$

Now, through Proposition 8.1 we can obtain the following upper bound.

**Theorem 8.2.** *For any  $k \geq 2$  and  $r > 0$ , the estimator  $\tilde{S}$  from (8.1) satisfies the inequality*

$$\limsup_{T \rightarrow \infty} T^{k_*} \sup_{\vartheta \in \Theta_{k,r}} v(\vartheta) \mathbf{E}_\vartheta \|\tilde{S} - S\|_n^2 \leq l_*. \quad (8.10)$$

Note that Lemma A.1 implies the following upper bound.

**Theorem 8.3.** *The quadratic risk for the estimating procedure  $\tilde{S}$  from (8.1) has the following asymptotic upper bound*

$$\limsup_{T \rightarrow \infty} T^{k_*} \sup_{\vartheta \in \Theta_{k,r}} v(\vartheta) \mathcal{R}(\tilde{S}, S) \leq l_*. \quad (8.11)$$

**Remark 8.1.** *It should be noted that the inequality (8.11) together with Theorem 6.1 mean that the estimator  $\tilde{S}$  defined by (8.1) is efficient. Unfortunately, we can't calculate this estimator, since it depends on unknown parameters  $k$ ,  $r$  and  $\mathbf{J}_\vartheta$ . But, this estimator belongs to the family  $(\hat{S}_\lambda)_{\lambda \in \Lambda}$  with the weight vectors defined in (5.23). Therefore, through the oracle inequality we can estimate the risk of the model selection procedure with the risk of the estimator (5.23) and, therefore, using the property (8.11) we can provide the efficiency property for the procedure (5.17).*

## 9 Proofs

### 9.1 Proof of Theorem 6.1

First, we introduce the following auxiliary class

$$\Theta_{k,r}^0 = \{\vartheta = (S, b_0) : S \in W_{k,r}, \quad b_0 \equiv 1\}.$$

It is clear that  $\Theta_{k,r}^0 \subset \Theta_{k,r}$  and

$$\sup_{\vartheta \in \Theta_{k,r}^0} v(\vartheta) \mathcal{R}(\tilde{S}, S) \geq \sup_{\vartheta \in \Theta_{k,r}} v(\vartheta) \mathcal{R}(\tilde{S}, S).$$

Using here Theorem 5.2 from [17] we obtain Theorem 6.1.  $\square$

### 9.2 Proof of Theorem 8.2

First note that, in view of Proposition 7.2, to prove this theorem it suffices to show that

$$\limsup_{T \rightarrow \infty} T^{k_*} \sup_{\vartheta \in \Theta_{k,r}} v(\vartheta) \mathbf{E}_\vartheta \chi_{\mathbf{G}_*} \|\tilde{S} - S\|_n^2 \leq l_* . \quad (9.1)$$

Indeed, note that due to (5.3)-(5.6) on the set  $\mathbf{G}_*$  we obtain that

$$\begin{aligned} \|\tilde{S} - S\|_n^2 &= \sum_{j=1}^n \left( \tilde{\lambda}^2(j) (\theta_{j,n} + \zeta_{j,n})^2 - 2\tilde{\lambda}(j) (\theta_{j,n} + \zeta_{j,n}) \theta_{j,n} + \theta_{j,n}^2 \right) \\ &= \sum_{j=1}^n \left( (\tilde{\lambda}(j) - 1)^2 \theta_{j,n}^2 + 2\tilde{\lambda}(j) (\tilde{\lambda}(j) - 1) \zeta_{j,n} \theta_{j,n} + \tilde{\lambda}^2(j) \zeta_{j,n}^2 \right), \end{aligned} \quad (9.2)$$

where  $\theta_{j,n} = (S, \phi_j)_n$  and  $\tilde{\lambda}$  is defined in (8.1). Taking into account that

$$\zeta_{j,n} = \mathbf{g}_{j,n} + \sqrt{\frac{\mathbf{x}_1 - \mathbf{x}_0}{n}} \xi_{j,n},$$

using the inequality

$$2xy \leq \epsilon x^2 + \epsilon^{-1} y^2,$$

we have from (9.2) on the set  $\mathbf{G}_*$

$$\begin{aligned} \|\tilde{S} - S\|_n^2 &\leq (1 + \epsilon) \left( \sum_{j=1}^n (1 - \tilde{\lambda}(j))^2 \theta_{j,n}^2 + B_1 \right) + \left( 1 + \frac{2}{\epsilon} \right) B_2 \\ &\quad + 2\sqrt{\frac{\mathbf{x}_1 - \mathbf{x}_0}{n}} \sum_{j=1}^n \tilde{\lambda}(j) (\tilde{\lambda}(j) - 1) \theta_{j,n} \xi_{j,n}, \end{aligned} \quad (9.3)$$

where

$$B_1 = \frac{\mathbf{x}_1 - \mathbf{x}_0}{n} \sum_{j=1}^n \tilde{\lambda}^2(j) \xi_{j,n}^2 \quad \text{and} \quad B_2 = \sum_{j=1}^n \tilde{\lambda}^2(j) \mathbf{g}_{j,n}^2. \quad (9.4)$$

Moreover, taking into account the property  $\mathbf{E} \xi_{j,n} = 0$ , one gets

$$\begin{aligned} \mathbf{E}_\vartheta \|\tilde{S} - S\|_n^2 \chi_{\mathbf{G}_*} &\leq (1 + \epsilon) \left( \sum_{j=1}^n (1 - \tilde{\lambda}(j))^2 \theta_{j,n}^2 + \mathbf{E}_\vartheta B_1 \right) \\ &\quad + \left( 1 + \frac{2}{\epsilon} \right) \mathbf{E}_\vartheta B_2 \chi_{\mathbf{G}_*} - 2\mathbf{E}_\vartheta B_3 \chi_{\mathbf{G}_*^c}, \end{aligned} \quad (9.5)$$

where

$$B_3 = \sqrt{\frac{\mathbf{x}_1 - \mathbf{x}_0}{n}} \sum_{j=1}^n \tilde{\lambda}(j) (\tilde{\lambda}(j) - 1) \xi_{j,n} \theta_{j,n}.$$

Note that

$$(\mathbf{x}_1 - \mathbf{x}_0) \mathbf{E}_\vartheta \left( \xi_{j,n}^2 | \mathcal{G}_{N_0} \right) = \frac{\mathbf{s}_n}{n} + \frac{\mathbf{x}_1 - \mathbf{x}_0}{n} \sum_{l=1}^n \sigma_l^2 \bar{\phi}_j(z_l) \quad (9.6)$$

where  $\mathbf{s}_n$  is defined in (7.2) and  $\bar{\phi}_j(z_l) = (\mathbf{x}_1 - \mathbf{x}_0) \phi_j^2(z_l) - 1$ . Setting now

$$\bar{\mathbf{s}}_n = \mathbf{s}_n - \frac{\mathbf{J}_\vartheta}{\mathbf{x}_1 - \mathbf{x}_0},$$

we represent the terme  $\mathbf{E}_\vartheta B_1$  as

$$\begin{aligned} \mathbf{E}_\vartheta B_1 &= \frac{1}{n^2} \mathbf{E}_\vartheta \sum_{j=1}^n \tilde{\lambda}^2(j) \left( \mathbf{s}_n + (\mathbf{x}_1 - \mathbf{x}_0) \sum_{l=1}^n \sigma_l^2 \bar{\phi}_j(x_l) \right) \\ &= \frac{\mathbf{J}_\vartheta}{(\mathbf{x}_1 - \mathbf{x}_0) n^2} \sum_{j=1}^n \tilde{\lambda}^2(j) + B_{11} + B_{12}, \end{aligned} \quad (9.7)$$

where

$$B_{11} = \frac{1}{n^2} \sum_{j=1}^n \tilde{\lambda}^2(j) \mathbf{E}_\vartheta \bar{\mathbf{s}}_n, \quad B_{12} = \frac{\mathbf{x}_1 - \mathbf{x}_0}{n^2} \mathbf{E}_\vartheta \sum_{j=1}^n \tilde{\lambda}^2(j) \sum_{l=1}^n \sigma_l^2 \bar{\phi}_j(x_l).$$

Now from (9.5) and (9.7) we obtain that

$$\begin{aligned} \mathbf{E}_\vartheta \|\tilde{S} - S\|_n^2 \chi_{\mathbf{G}_*} &\leq (1 + \epsilon) \left( \tilde{\Upsilon}_T + B_{11} + B_{12} \right) \\ &\quad + \left( 1 + \frac{2}{\epsilon} \right) \mathbf{E}_\vartheta B_2 \chi_{\mathbf{G}_*} - 2\mathbf{E}_\vartheta B_3 \chi_{\mathbf{G}_*^c}. \end{aligned} \quad (9.8)$$



The property (8.9) and Proposition 7.3 yield for the term  $B_{11}$

$$\limsup_{T \rightarrow \infty} T^{k_*} \sup_{\vartheta \in \Theta_{k,r}} |B_{11}| = 0.$$

As to the term  $B_{12}$ , due to Lemma A.4, one has

$$\begin{aligned} |B_{12}| &\leq \frac{\mathbf{x}_1 - \mathbf{x}_0}{n^2} \mathbf{E}_\vartheta \sum_{l=1}^n \sigma_l^2 \left| \sum_{j=1}^n \tilde{\lambda}^2(j) \bar{\phi}_j(x_l) \right| \\ &\leq 2^{k+1} \frac{\mathbf{x}_1 - \mathbf{x}_0}{n^2} \mathbf{E}_\vartheta \sum_{l=1}^n \sigma_l^2 = 2^{k+1} \mathbf{E}_\vartheta \frac{\mathbf{S}_n}{n^2}. \end{aligned} \quad (9.9)$$

Therefore, in view of Proposition 7.3, we obtain that

$$\limsup_{T \rightarrow \infty} T^{k_*} \sup_{\vartheta \in \Theta_{k,r}} |B_{12}| = 0.$$

Moreover, by the definition,

$$B_2 = \sum_{j=1}^n \tilde{\lambda}^2(j) \mathbf{g}_{j,n}^2 \quad \text{and} \quad \mathbf{g}_{j,n} = (\mathbf{g}, \phi_j)_n.$$

Therefore,

$$B_2 \leq \sum_{j=1}^n \mathbf{g}_{j,n}^2 = \|\mathbf{g}\|_n^2$$

and

$$\mathbf{E}_\vartheta \mathbf{1}_{\mathbf{G}_*} B_2 \leq \mathbf{E}_\vartheta \mathbf{1}_{\mathbf{G}_*} \|\mathbf{g}\|_n^2 = \frac{\mathbf{x}_1 - \mathbf{x}_0}{n} \sum_{k=1}^n \mathbf{E}_\vartheta \mathbf{1}_{\mathbf{G}_*} \mathbf{g}_k^2.$$

It is easy to verify (see [21]) that

$$\sup_{\vartheta \in \Theta} \max_{1 \leq k \leq n} \mathbf{E}_\vartheta \mathbf{1}_{\mathbf{G}_*} \mathbf{g}_k^2 \leq M^2 h^2 + o(1/T).$$

These bounds yield that

$$\lim_{T \rightarrow \infty} T^{\frac{2k}{2k+1}} \sup_{\vartheta \in \Theta_{k,r}} \mathbf{E}_\vartheta \mathbf{1}_{\mathbf{G}_*} B_2 = 0. \quad (9.10)$$

To estimate the last term in (9.5) note that through Lemma A.5 we can estimate the square of the expectation of this term as

$$\mathbf{E}_\vartheta B_3^2 \leq \sigma_{1,*} \frac{\mathbf{x}_1 - \mathbf{x}_0}{n} \sum_{j=1}^n \theta_{j,n}^2 = \frac{\sigma_{1,*} (\mathbf{x}_1 - \mathbf{x}_0)}{n} \|S\|_n^2.$$

Therefore,

$$\mathbf{E}_\vartheta \mathbf{1}_{\mathbf{G}_*^c} |B_3| \leq \sqrt{\frac{\sigma_{1,*}(\mathbf{x}_1 - \mathbf{x}_0)}{n}} \|S\|_n \mathbf{P}_\vartheta(\mathbf{G}_*^c)$$

and taking into account Proposition 7.2, we obtain that

$$\lim_{T \rightarrow \infty} T^{\frac{2k}{2k+1}} \sup_{\vartheta \in \Theta_{k,r}} \mathbf{E}_\vartheta \mathbf{1}_{\mathbf{G}_*^c} |B_3| = 0.$$

Therefore, using Proposition 8.1 in (9.8) we obtain Theorem 8.2.  $\square$

### 9.3 Proof of Theorem 6.2

Taking into account that for sufficiently large  $T$  the estimator (8.1) belongs to the family  $(\widehat{S}_\lambda)_{\lambda \in \Lambda}$  indexed by the set (5.23), we obtain that

$$\limsup_{T \rightarrow \infty} T^{k_*} \mathbf{E}_\vartheta \|\widehat{S}_* - S\|^2 \leq \limsup_{T \rightarrow \infty} T^{k_*} \mathbf{E}_\vartheta \|\widehat{S}_\lambda - S\|^2.$$

So, Theorem 8.2 implies immediately Theorem 6.2.  $\square$

**Acknowledgments.** The work of the last author was partially supported by the Russian Federal Professor program, project no. 1.472.2016/1.4 (the Ministry of Science and Higher Education of the Russian Federation).

## A Appendix

### A.1 Properties of the trigonometric basis.

**Lemma A.1.** *Let  $f$  be an absolutely continuous  $[\mathbf{x}_0, \mathbf{x}_1] \rightarrow \mathbb{R}$  function with  $\|f\| < \infty$  and  $g$  be  $[\mathbf{x}_0, \mathbf{x}_1] \rightarrow \mathbb{R}$  a step-wise function of the form*

$$g(z) = \sum_{j=1}^n c_j \chi_{(z_{j-1}, z_j]}(z),$$

where  $c_j$  are some constants and the sequence  $(z_j)_{0 \leq j \leq n}$  is given in (3.1). Then, for any  $\tilde{\varepsilon} > 0$ , the function  $\Delta = f - g$  satisfies the following inequalities

$$\|\Delta\|^2 \leq (1 + \tilde{\varepsilon}) \|\Delta\|_n^2 + \left(1 + \frac{1}{\tilde{\varepsilon}}\right) \frac{\|\dot{f}\|^2}{n^2} (\mathbf{x}_1 - \mathbf{x}_0)^2,$$

and

$$\|\Delta\|_n^2 \leq (1 + \tilde{\varepsilon}) \|\Delta\|^2 + \left(1 + \frac{1}{\tilde{\varepsilon}}\right) \frac{\|\dot{f}\|^2}{n^2} (\mathbf{x}_1 - \mathbf{x}_0)^2.$$

The proof is given in Lemma A.2 from [31].

**Lemma A.2.** *For any  $1 \leq j \leq n$ , the trigonometric Fourier coefficients  $(\theta_{j,n})_{1 \leq j \leq n}$  of the functions  $S$  from the class  $W_{k,r}$  satisfy, for any  $\tilde{\varepsilon} > 0$ , the following inequality*

$$\theta_{j,n}^2 \leq (1 + \tilde{\varepsilon}) \theta_j^2 + (1 + \tilde{\varepsilon}^{-1}) \frac{\check{r}_k}{n^{2k}}, \quad (\text{A.1})$$

where

$$\check{r}_k = \frac{2r(\pi^2 + 1)(\mathbf{x}_1 - \mathbf{x}_0)^{2k}}{\pi^{2k}}.$$

**Proof.** First we represent the function  $S$  as the Fourier series in  $\mathcal{L}_2([\mathbf{x}_0, \mathbf{x}_1])$  :

$$S(x) = \sum_{l=1}^n \theta_l \phi_l(x) + \Delta_n(x),$$

where

$$\Delta_n(x) = \sum_{l>n} \theta_l \phi_l(x). \quad (\text{A.2})$$

Making now the Fourier decomposition of  $S$  on the grid  $(z_k)_{1 \leq k \leq n}$  with an odd  $n$ , one gets the Fourier coefficients :

$$\theta_{j,n} = (S, \phi_j)_n = \theta_j + (\Delta_n, \phi_j)_n$$

and, for any  $0 < \tilde{\varepsilon} < 1$ ,

$$\theta_{j,n}^2 \leq (1 + \tilde{\varepsilon}) \theta_j^2 + (1 + \tilde{\varepsilon}^{-1}) \|\Delta_n\|_n^2.$$

By applying Lemma A.1 with  $f = \Delta_n, g = 0$ , we obtain that

$$\begin{aligned} \|\Delta_n\|_n^2 &\leq 2\|\Delta_n\|^2 + 2 \frac{\|\dot{\Delta}_n\|^2 (\mathbf{x}_1 - \mathbf{x}_0)^2}{n^2} = 2 \sum_{l>n} \theta_l^2 + 2 \frac{\|\dot{\Delta}_n\|^2 (\mathbf{x}_1 - \mathbf{x}_0)^2}{n^2} \\ &\leq 2ra_{n+1}^{-1} + 2 \frac{\|\dot{\Delta}_n\|^2 (\mathbf{x}_1 - \mathbf{x}_0)^2}{n^2}. \end{aligned}$$

Taking account of (6.3) and that  $2[l/2] \geq l - 1$  for  $l \geq 2$ , we have

$$\|\Delta_n\|_n^2 \leq \frac{2r(\mathbf{x}_1 - \mathbf{x}_0)^{2k}}{\pi^{2k} n^{2k}} + 2 \frac{\|\dot{\Delta}_n\|^2 (\mathbf{x}_1 - \mathbf{x}_0)^2}{n^2}.$$

Note here, that for any  $n \geq 1$ ,

$$\begin{aligned} \|\dot{\Delta}_n\|^2 &= \frac{(2\pi)^2}{(\mathbf{x}_1 - \mathbf{x}_0)^2} \sum_{l>n} \theta_l^2 [l/2]^2 = \frac{(\mathbf{x}_1 - \mathbf{x}_0)^{2(k-1)}}{\pi^{2(k-1)}} \sum_{l>n} \frac{a_l \theta_l^2}{(2[l/2])^{2(k-1)}} \\ &\leq \frac{(\mathbf{x}_1 - \mathbf{x}_0)^{2(k-1)}}{\pi^{2(k-1)}} \sum_{l>n} \frac{a_l \theta_l^2}{(l-1)^{2(k-1)}} \leq \frac{r(\mathbf{x}_1 - \mathbf{x}_0)^{2(k-1)}}{\pi^{2(k-1)} n^{2(k-1)}}. \end{aligned} \quad (\text{A.3})$$

Hence Lemma A.2  $\square$

**Lemma A.3.** *For any  $n \geq 2$ ,  $1 \leq m < n$  and  $r > 0$ , the coefficients  $(\theta_{j,n})_{1 \leq j \leq n}$  of functions  $S$  from the class  $W_{k,r}$  satisfy, for any  $\tilde{\varepsilon} > 0$ , the following inequality*

$$\sum_{j=m+1}^n \theta_{j,n}^2 \leq (1 + \tilde{\varepsilon}) \sum_{j \geq m+1} \theta_j^2 + (1 + \tilde{\varepsilon}^{-1}) \frac{\tilde{r}_1}{n^2 m^{2(k-1)}}, \quad (\text{A.4})$$

where

$$\tilde{r}_1 = \frac{r(\mathbf{x}_1 - \mathbf{x}_0)^{2k}}{\pi^{2(k-1)}}.$$

**Proof.** First we note that

$$\sum_{j=m+1}^n \theta_{j,n}^2 = \min_{x_1, \dots, x_m} \|S - \sum_{j=1}^m x_j \phi_j\|_n^2 \leq \|\Delta_m\|_n^2,$$

where the function  $\Delta_m(\cdot)$  is defined in (A.2). By applying Lemma A.1 with  $f = \Delta_m, g = 0$ , and taking account of the inequality (A.3), we obtain the bound (A.4). Hence Lemma A.3  $\square$

**Lemma A.4.** *For any  $k \geq 1$ ,*

$$\sup_{n \geq 2} n^{-k} \sup_{x \in [\mathbf{x}_0, \mathbf{x}_1]} \left| \sum_{l=2}^n l^k \bar{\phi}_l(x) \right| \leq 2^k, \quad (\text{A.5})$$

where  $\bar{\phi}_l(x) = (\mathbf{x}_1 - \mathbf{x}_0) \phi_l^2(x) - 1$ .

Proof of this result is given in Lemma A.2 from [15].

## A.2 Technical lemma

**Lemma A.5.** *Let  $(u_j)_{1 \leq j \leq n}$  be nonrandom sequence. Then*

$$\mathbf{E}_\vartheta \left( \sum_{j=1}^n u_j \xi_{j,n} \right)^2 \leq \sigma_{1,*} \sum_{j=1}^n u_j^2.$$

**Proof.** First we denote  $D_n = \sum_{j=1}^n u_j \xi_{j,n}$ . Now, using the definition of  $\xi_{j,n}$  in (5.5) we obtain that

$$D_n = \sqrt{\frac{\mathbf{x}_1 - \mathbf{x}_0}{n}} \sum_{l=1}^n \sigma_l \xi_{l,n} \sum_{j=1}^n u_j \phi_j(z_l).$$

This implies

$$\begin{aligned} \mathbf{E}_\vartheta D^2(u) &= \frac{\mathbf{x}_1 - \mathbf{x}_0}{n} \mathbf{E}_\vartheta \sum_{l=1}^n \sigma_l^2 \left( \sum_{j=1}^n u_j \phi_j(z_l) \right)^2 \\ &\leq \sigma_{1,*} \sum_{j=1}^n u^2(j) = \sigma_{1,*} \|u\|^2. \end{aligned}$$

Hence Lemma 3.1.  $\square$

## References

- [1] Arato, M. (1982) *Linear stochastic systems with constant coefficients. A statistical approach*. Springer-Verlag, Berlin, Heidelberg, New York
- [2] Bayisa, F.L., Zhou, Z., Cronie, O. and Yu, J. (2019) Adaptive algorithm for sparse signal recovery. *Digital Signal Processing*, **87**, 10 – 18.
- [3] Comte, F., Genon-Catalot, V. and Rozenholc, Y. (2009) Nonparametric estimation for a discretely observed integrated diffusion model. *Stochastic Processes and their Applications*. **119** 811–834.
- [4] Dalalyan, A.S. (2005) Sharp Adaptive Estimation of the Drift Function for Ergodic Diffusion. *Annals of Statistics*, **33** (6), 2507– 2528.
- [5] Dalalyan, A.S., and Kutoyants, Yu.A. (2002) Asymptotically Efficient Trend Coefficient Estimation for Ergodic Diffusion. *Mathematical Methods of Statistics*, **11** (4), 402–427.

- [6] Florens-Zmirou, D. (1993) On estimating the diffusion coefficient from discret observations. *J. Appl. Probab.* **30**(4), p. 790–804.
- [7] Fujimori, K. (2019) The Danzing selector for a linear model of diffusion processes. - *Statistical Inference for Stochastic Processes*, **22**, p. 475 – 498.
- [8] De Gregorio, A., Iacus, S.M. (2012) Adaptive LASSO-type estimation for multivariate diffusion processes. *Econom. Theory*, **28** (4), p. 838 – 860.
- [9] Galtchouk, L.I. (1978) Existence and uniqueness of a solution for stochastic equations with respect to semimartingales. *Theory of Probability Appl.*, **23**, 751-763.
- [10] Galtchouk, L. and Pergamenshchikov, S. (2001) Sequential nonparametric adaptive estimation of the drift coefficient in diffusion processes. *Mathematical Methods of Statistics*, **10**, 3, p. 316-330.
- [11] Galtchouk, L. and Pergamenshchikov, S. (2004) Nonparametric sequential estimation of the drift in diffusion via model selection. *Mathematical Methods of Statistics*, **13**, 25-49.
- [12] Galtchouk, L. and Pergamenshchikov, S. (2005) Nonparametric sequential minimax estimation of the drift coefficient in diffusion processes. *Sequential Analysis*, **24**(3), 303-330.
- [13] Galtchouk, L. and Pergamenshchikov, S. (2006) Asymptotically efficient sequential kernel estimates of the drift coefficient in ergodic diffusion processes. *Statistical Inference for Stochastic Processes*, **9**, 1-16.
- [14] Galtchouk, L. and Pergamenshchikov, S. (2007) Uniform concentration inequality for ergodic diffusion processes. *Stochastic Processes and Applications*, **117**, 830-839.
- [15] Galtchouk, L. and Pergamenshchikov, S. (2009a) Sharp non-asymptotic oracle inequalities for nonparametric heteroscedastic regression models. *Journal of Nonparametric Statistics*, **21**, 1-16.
- [16] Galtchouk, L. and Pergamenshchikov, S. (2009b) Adaptive asymptotically efficient estimation in heteroscedastic nonparametric regression. *Journal of the Korean Statistical Society*, **38**(4), 305-322.

- [17] Galtchouk, L. and Pergamenshchikov, S. (2011) Adaptive sequential estimation for ergodic diffusion processes in quadratic metric. *Journal of Nonparametric Statistics*, **23**(2), 255-285.
- [18] Galtchouk, L. and Pergamenshchikov, S. (2013) Uniform concentration inequality for ergodic diffusion processes observed at discrete times. *Stochastic Processes and Applications*, **123**(1), 91-109.
- [19] Galtchouk, L. and Pergamenshchikov, S. (2014) Geometric ergodicity for classes of homogeneous Markov chains. - *Stochastic Processes and their Applications*, 2014, v. 124, 10, p. 3362-3391.
- [20] Galtchouk, L. and Pergamenshchikov, S. (2015) Efficient pointwise estimation based on discrete data in ergodic nonparametric diffusions. *Bernoulli*, **21**(4), 2569-2594.
- [21] Galtchouk, L. and Pergamenshchikov, S. (2019) Non asymptotic sharp oracle inequalities for high dimensional ergodic diffusion models. - Preprint, 2019. <https://hal.archives-ouvertes.fr/hal-02387034>
- [22] Gihman, I.I. and Skorohod, A.V. (1968) Stochastic differential equations. *Naukova Dumka*, Kiev.
- [23] Gobet, E., Hoffmann, M. and Reiss, M. (2004) Nonparametric estimation of scalar diffusions based on low frequency data. *Ann. Statist.*, **32**(5), 2223–2253.
- [24] Genon-Catalot, V. and Laredo, C. and Picard, D. (1992) Non-parametric estimation of the diffusion coefficient by wavelets methods. *Scand. J. Statist.* **19** 317–335.
- [25] Hoffmann, M. (1999) Adaptive estimation in diffusion processes. *Stochastic Processes and their Applications*, **79**, p. 135 - 163.
- [26] Ibragimov, I.A. and Hasminskii, R.Z. (1979) *Statistical Estimation: Asymptotic Theory*. Springer, Berlin, New York.
- [27] Jacod, J. (2000) Non-parametric kernel estimation of the coefficient of a diffusion. *Scand. J. Statist.* **27**(1) 83–96.
- [28] Kabanov, Yu. M. and Pergamenshchikov, S. M. (2003) *Two Scale Stochastic Systems: Asymptotic Analysis and Control*. - Applications of Mathematics, Stochastic Modelling and Applied Probability, 49, Springer-Verlag, Berlin, New York.

- [29] Karatzas, I. and Shreve, S.E. (1998) *Methods of Mathematical Finance*. Springer, New York.
- [30] Kutoyants, Yu.A. (2003) *Statistical Inferences for Ergodic Diffusion Processes*. Springer, Berlin.
- [31] Konev, V.V. and Pergamenshchikov, S.M. (2015) Robust model selection for a semimartingale continuous time regression from discrete data. *Stochastic processes and their applications*, **125**, 294 – 326.
- [32] Lamberton, D. and Lapeyre, B. (1996) *Introduction to stochastic calculus applied to finance*. Chapman & Hall, London.
- [33] Pinsker, M.S. (1981) Optimal filtration of square integrable signals in gaussian noise. *Problems Inform. Transmission*, **17**, p. 120-133.