



HAL
open science

Adaptive sequential estimation for ergodic diffusion processes in quadratic metric. Part 1: Sharp non-asymptotic oracle inequalities.

Leonid Galtchouk, Sergey Pergamenshchikov

► **To cite this version:**

Leonid Galtchouk, Sergey Pergamenshchikov. Adaptive sequential estimation for ergodic diffusion processes in quadratic metric. Part 1: Sharp non-asymptotic oracle inequalities.. 2007. hal-00177875

HAL Id: hal-00177875

<https://hal.science/hal-00177875>

Preprint submitted on 9 Oct 2007

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.

Adaptive sequential estimation for ergodic diffusion processes in quadratic metric.

Part 1: Sharp non-asymptotic oracle inequalities

By Leonid Galtchouk and Sergey Pergamenshchikov *

University of Strasbourg and University of Rouen

Abstract

An adaptive nonparametric estimation procedure is constructed for the estimation problem of the drift coefficient in diffusion processes. A non-asymptotic oracle upper bound for the quadratic risk is obtained.

1 2

*The second author is partially supported by the RFFI-Grant 04-01-00855.

¹AMS 2000 Subject Classification : primary 62G08; secondary 62G05, 62G20

²Key words: adaptive estimation, asymptotic bounds, efficient estimation, nonparametric estimation, non-asymptotic estimation, oracle inequality, Pinsker's constant.

1 Introduction

Let $(\Omega, \mathcal{F}, (\mathcal{F})_{t \geq 0}, \mathbf{P})$ be a filtered probability space on which the following stochastic differential equation is defined

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

where $(w_t)_{t \geq 0}$ is a scalar standard Wiener process, the initial value y_0 is a given constant and $S(\cdot)$ is an unknown function.

The problem is to estimate the function S from observations of $(y_t)_{0 \leq t \leq T}$ in non-asymptotic setting, and to obtain sharp non-asymptotic bounds for a minimax quadratic risk. This problem is important for applications.

It seems that, for the first time, the problem of non-asymptotic parameter estimation for diffusion processes has been studied in [1] for wobbling analysis of the axis of the equator. There, for a special diffusion process, the exact distribution of the ML-estimators of unknown parameters has been obtained for any finite sample time T . Unfortunately, in the majority cases, when the sample time is finite, it is difficult to study classical estimators such as LS-estimators or ML-estimators since they are non-linear functionals of observations.

The paper [18] has shown that many difficulties in non-asymptotic parameter estimation for one-dimensional diffusion processes can be overcome by the sequential approach. It turns out that the sequential ML-estimator is more advantageous than the usual one. In particular, it is possible to calculate non-asymptotic bounds for quadratic risk in the sequential procedure. By making use of the sequential approach non-asymptotic parameter estimation problems have been studied in [16], [5] for multidimensional diffusion processes and recently in [6] for multidimensional continuous and discrete

time semimartingales. In the paper [17] a truncated sequential method has been developed for parameter estimation in diffusion processes.

The sequential approach to nonparametric minimax estimation problem of the drift coefficient in ergodic diffusion processes has been developed in [7]-[10],[12]. The papers [7], [8],[10] deal with sequential pointwise kernel estimators of the drift coefficient. A non-asymptotic upper bound was obtained for absolute error risks. The estimators yield also the optimal convergence rate as the sample time $T \rightarrow \infty$. In the paper [8] it is shown that this procedure is minimax and adaptive in the both cases when either the smoothness is known or unknown. The same type of the sequential kernel estimators is used in the paper [9] for the nonparametric estimation in the L_2 -metric of the drift coefficient via model selection. A non-asymptotic upper bound for the quadratic risk is proved. The procedure is minimax and adaptive in the asymptotic setting as well. A sequential asymptotically efficient kernel estimator is constructed for pointwise drift estimation in [12].

A new method using the concentration inequality for the nonlinear LS-estimators analysis has been proposed in [3], [19] for homoscedastic nonparametric gaussian regression. A non-asymptotic upper bound for a quadratic risk (the Oracle inequality) was obtained which is best over the LS-estimators. This method was developed in [4] for any estimators to regression models with depending noises for which a non-asymptotic Oracle inequality has been proved. Further development of this method was done recently in [11] for heteroscedastic nonparametric regression models with a varying variance depending on unknown regression function. In the last paper weighted LS-estimators were used to obtain the non-asymptotic Oracle inequality.

This paper deals at the first time with the non-asymptotic estimation of the drift coefficient S in adaptive setting for the quadratic risk

$$\mathcal{R}(\hat{S}_T, S) = \mathbf{E}_S \|\hat{S}_T - S\|^2, \quad (1.2)$$

where \hat{S}_T is an estimator of S , $\|S\|^2 = \int_a^b S^2(x)dx$, a, b are some real numbers, $b - a \geq 1$. Here \mathbf{E}_S is the expectation with respect to the distribution law \mathbf{P}_S of the process $(y_t)_{t \geq 0}$ given the drift S .

Our main goal is to construct by sequential methods an adaptive estimator \hat{S}_* of the function S in (1.1) and to show that the quadratic risk of this estimator satisfies a sharp non-asymptotic oracle inequality with respect to a family of weighted LS-estimators.

The sharp non-asymptotic oracle inequality with respect to a estimators family $(\hat{S}_\alpha, \alpha \in \mathcal{A})$ means that the principal term in the upper bound for the quadratic risk has a coefficient which closes to one, i.e.

$$\mathcal{R}(\hat{S}_T, S) \leq (1 + D(\rho)) \min_{\alpha \in \mathcal{A}} \mathcal{R}(\hat{S}_\alpha, S) + \frac{\mathcal{B}_T(\rho)}{T}, \quad (1.3)$$

where the function $D(\rho) \rightarrow 0$ as $\rho \rightarrow 0$ and $\mathcal{B}_T(\rho)$ is some slowly varying function, i.e. for any $\delta > 0$ and $\rho > 0$,

$$\lim_{T \rightarrow \infty} \frac{\mathcal{B}_T(\rho)}{T^\delta} = 0. \quad (1.4)$$

The goal of this paper is to obtain this inequality for model (1.1). To this end we pass to a discrete time regression model by making use of the truncated sequential procedure introduced in [8], [9] and [12] and we come to the Γ - regression scheme, i.e. the sequential kernel estimator at the point x_k on the set Γ has the following representation

$$Y_k = S(x_k) + \zeta_k. \quad (1.5)$$

Here, see [9], in the contrast with the classical regression model, the noise sequence $(\zeta_k)_{k \geq 1}$ has a complex structure, i.e.

$$\zeta_k = \sigma_k \xi_k + \delta_k, \quad (1.6)$$

where $(\sigma_k)_{k \geq 1}$ is a sequence of observed random variables, $(\delta_k)_{k \geq 1}$ is a sequence of bounded random variables and $(\xi_k)_{k \geq 1}$ is a sequence of i.i.d. random variables $\mathcal{N}(0, 1)$ which are independent of $(\sigma_k)_{k \geq 1}$.

In this paper we propose a new method to construct adaptive estimators that is based on the approach proposed in [11] and on the penalisation method proposed in [2]. In our method we propose to add an penalty term with a small coefficient into the quadratic loss functional used to find an estimator. This new functional criterion with the penalty term allows us to obtain a sharp oracle inequality and the above small coefficient provides that the principal term of the oracle inequality has a coefficient closes to one. The new criterion allows us to overcome difficulties bound up with depending noises appearing after the discrete Fourier transformation to model (1.5).

To estimate the function S in model (1.5) we make use of the estimator family $(\hat{S}_\alpha, \alpha \in \mathcal{A})$, where \hat{S}_α is a weighted least squares estimator with the Pinsker weights. For this family similarly to [11], we construct a special selection rule, i.e. a random variable α_* with values in \mathcal{A} , for which we define the selection estimator as $\hat{S}_* = \hat{S}_{\alpha_*}$. We show (see Theorem 3.2) that this selection rule for any $0 < \rho < 1/6$ and any $T \geq 32$ satisfies the oracle inequality (1.3) with

$$D(\rho) = \frac{11\rho + 7\rho^2 + 3\rho^3}{1 - 6\rho} \quad (1.7)$$

and with some function $\mathcal{B}_T(\rho)$ satisfying property (1.4).

We make use of the non-asymptotic analysis method from [11] based on the non-asymptotic investigation proposed in [2] for the family of least-squares estimators and developed in [4] for some other families of estimators.

In the second part of this investigation we prove the asymptotic efficiency of the proposed procedure that is we show that the oracle inequality (1.3) provides the minimal quadratic risk which coincides with the Pinsker constant. Juste in this sense the inequality (1.3) is sharp.

The paper is organized as follows. In the next Section we give the definitions of functional classes and the passage from diffusion model (1.1) to a discrete time regression model. In Section 3 the sharp non-asymptotic oracle inequality is given for the regression model (Theorem 3.1). In Section 4 we prove Theorem 3.1 and Theorem 3.2. Appendix contains the proofs of some technique results.

2 Passage to a discrete time regression model

To obtain a good estimate of the function S , it is necessary to impose some conditions on the function S 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) in (1.1) returns to any vicinity of each point $x \in [a, b]$ infinite times. The ergodicity of (y_t) provides this property.

Let $L > 1$. We define the following functional class :

$$\begin{aligned} \Sigma_L = \{S \in Lip_L(\mathbb{R}) : \sup_{|z| \leq L} |S(z)| \leq L; \forall |x| \geq L, \\ \exists \dot{S}(x) \in \mathcal{C} \text{ such that } -L \leq \dot{S}(x) \leq -1/L\}, \end{aligned} \quad (2.1)$$

where

$$Lip_L(\mathbb{R}) = \left\{ f \in \mathcal{C}(\mathbb{R}) : \sup_{x,y \in \mathbb{R}} \frac{|f(x) - f(y)|}{|x - y|} \leq L \right\}.$$

First of all, note that if $S \in \Sigma_L$, then there exists the ergodic density

$$q(x) = q_S(x) = \frac{\exp\{2 \int_0^x S(z) dz\}}{\int_{-\infty}^{+\infty} \exp\{2 \int_0^y S(z) dz\} dy} \quad (2.2)$$

(see, e.g., Gihman and Skorohod (1972), Ch.4, 18, Th2). It is easy to see that this density is uniformly bounded in the class (2.1), i.e.

$$q^* = \sup_{x \in \mathbb{R}} \sup_{S \in \Sigma_L} q_S(x) < +\infty$$

and bounded away from zero on the interval $[-b_*, b_*]$, i.e.

$$q_* = \inf_{|x| \leq b_*} \inf_{S \in \Sigma_L} q_S(x) > 0, \quad (2.3)$$

where $b_* = 1 + |a| + |b|$. Moreover, we note that the functions from Σ_L are uniformly bounded, i.e.

$$s_* = \sup_{a \leq x \leq b} \sup_{S \in \Sigma_L} S^2(x) < \infty. \quad (2.4)$$

To obtain the oracle inequality we pass to discrete time regression model (1.5) by the same way as in [9]. We start with the partition of the interval $[a, b]$ by points $(x_k)_{1 \leq k \leq n}$ defined as

$$x_k = a + \frac{k}{n}(b - a) \quad (2.5)$$

with $n = 2[(T - 1)/2] + 1$ ($[x]$ is the integer part of x). At each point x_k we will estimate the function S by the sequential kernel estimator from [9]-[10].

We fix some $0 < t_0 < T$ and we set

$$\tau_k = \inf\{t \geq t_0 : \int_{t_0}^t Q\left(\frac{y_s - x_k}{h}\right) ds \geq H_k\} \quad (2.6)$$

and

$$S_k^* = \frac{1}{H_k} \int_{t_0}^{\tau_k} Q\left(\frac{y_s - x_k}{h}\right) dy_s, \quad (2.7)$$

where $Q(z) = \mathbf{1}_{\{|z| \leq 1\}}$, $h = (b-a)/(2n)$ and H_k is a positive threshold. From (1.1) it is easy to obtain that

$$S_k^* = S(x_k) + \zeta_k.$$

The error term ζ_k is represented as the following sum of the approximated term B_k and the stochastic term:

$$\zeta_k = B_k + \frac{1}{\sqrt{H_k}} \xi_k,$$

where

$$\begin{aligned} B_k &= \frac{1}{H_k} \int_{t_0}^{\tau_k} Q\left(\frac{y_s - x_k}{h}\right) (S(y_s) - S(x_k)) ds, \\ \xi_k &= \frac{1}{\sqrt{H_k}} \int_{t_0}^{\tau_k} Q\left(\frac{y_s - x_k}{h}\right) dw_s. \end{aligned}$$

Taking into account that the function S is lipschitzian, we obtain the upper bound for the approximated term as

$$|B_k| \leq L h. \quad (2.8)$$

It is easy to see that the random variables $(\xi_k)_{1 \leq k \leq n}$ are i.i.d. normal $\mathcal{N}(0, 1)$.

Moreover, in [12] it is established that the efficient kernel estimator of type (2.7) has the stochastic term distributed as $\mathcal{N}(0, 2Thq_S(x_k))$. Therefore for the efficient estimation at each point x_k by kernel estimator (2.7) we need to estimate the ergodic density (2.2) from the observations $(y_t)_{0 \leq t \leq t_0}$. We put

$$\tilde{q}_T(x_k) = \max\{\hat{q}(x_k), \epsilon_T\},$$

where ϵ_T is positive, $0 < \epsilon_T < 1$,

$$\hat{q}(x_k) = \frac{1}{2t_0h} \int_0^{t_0} Q\left(\frac{y_s - x_k}{h}\right) ds. \quad (2.9)$$

Now we choose the threshold H_k in (2.6)–(2.7) as

$$H_k = (T - t_0)(2\tilde{q}_T(x_k) - \epsilon_T^2)h.$$

We suppose that for any $T \geq 32$ the parameters $t_0 = t_0(T)$ and ϵ_T satisfy the following conditions

$$16 \leq t_0 \leq \frac{T}{2} \quad \text{and} \quad \sqrt{2}t_0^{-1/8} \leq \epsilon_T \leq 1. \quad (2.10)$$

Moreover,

$$\lim_{T \rightarrow \infty} \epsilon_T = 0, \quad \lim_{T \rightarrow \infty} \frac{T\epsilon_T}{t_0(T)} = \infty. \quad (2.11)$$

Finally, we assume also for any $\delta > 0$ and $\mu > 0$,

$$\lim_{T \rightarrow \infty} T^\mu e^{-\delta\sqrt{t_0}} = 0. \quad (2.12)$$

For example, one can take, for $T \geq 32$,

$$t_0 = \max\{\min\{\ln^4 T, T/2\}, 16\} \quad \text{and} \quad \epsilon_T = \sqrt{2}t_0^{-1/8}.$$

We set now

$$\Gamma = \left\{ \max_{1 \leq k \leq n} \tau_k \leq T \right\} \quad \text{and} \quad Y_k = S_k^* \mathbf{1}_\Gamma. \quad (2.13)$$

Therefore on the set Γ one has the discrete time regression model (1.5) with

$$\delta_l = B_l \quad \text{and} \quad \sigma_l^2 = \frac{n}{(T - t_0)(\tilde{q}_T(x_l) - \epsilon_T^2/2)(b - a)}. \quad (2.14)$$

To study properties of the set Γ we introduce the following functions

$$\varrho(c, L) = 2c \left(e^{L^3} + \frac{2q^*}{q_*} + q^* \int_0^\infty e^{-z^2/L+2Lz} dz \right),$$

$$K_*(L) = L \max(y_0^2, L^2(1 + 2L^2 + 2L)^2) + y_0^2 + 1$$

and

$$\gamma(c, L) = \frac{1}{54\rho^2(c, L)K_*(L)}. \quad (2.15)$$

In Appendix A.1 we show the following result.

Proposition 2.1. *Suppose that the parameters t_0 and ϵ_T satisfy the conditions (2.10)–(2.12). Then*

$$\sup_{S \in \Sigma_L} \mathbf{P}(\Gamma^c) \leq T \mathbf{1}_{\{q_* < 2\epsilon_T\}} + 16Te^{-\gamma_1 \sqrt{t_0}} \mathbf{1}_{\{q_* \geq 2\epsilon_T\}} := \Pi_T \quad (2.16)$$

with

$$\gamma_1 = \gamma(2, L) \min(q_*^2, 1/8).$$

Note that the conditions (2.10)–(2.12) imply directly that for any $m > 0$

$$\lim_{T \rightarrow \infty} T^m \Pi_T = 0. \quad (2.17)$$

3 Oracle inequalities

In this section we consider the estimation problem for the following discrete time regression model. We suppose that on some set $\Gamma \subseteq \Omega$ we have the following nonparametric regression

$$Y_l = S(x_l) + \zeta_l, \quad 1 \leq l \leq n, \quad (3.1)$$

where $\zeta_l = \sigma_l \xi_l + \delta_l$ and n is odd. The points $(x_l)_{1 \leq l \leq n}$ are defined in (2.5) with $b - a \geq 1$. The function S is unknown and to be estimated from observations Y_1, \dots, Y_n . The quality of an estimator \hat{S} will be measured by the empiric squared error

$$\|\hat{S} - S\|_n^2 = (\hat{S} - S, \hat{S} - S)_n = \frac{b-a}{n} \sum_{l=1}^n (\hat{S}(x_l) - S(x_l))^2.$$

We suppose that $(\xi_i)_{1 \leq i \leq n}$ is a sequence of i.i.d. random variables $\mathcal{N}(0, 1)$, $(\sigma_l)_{1 \leq l \leq n}$ is a sequence of observed positive random variables independent of $(\xi_i)_{1 \leq i \leq n}$ and bounded by some nonrandom constant $\varsigma_n \geq 1$, i.e.

$$\max_{1 \leq l \leq n} \sigma_l^2 \leq \varsigma_n. \quad (3.2)$$

Concerning the random sequence $\delta = (\delta_l)_{1 \leq l \leq n}$ we suppose that

$$\mathbf{E}_S \|\delta\|_n^2 < \infty. \quad (3.3)$$

Note that the model (1.5) is a particular case of (3.1) with Γ defined in (2.13) and $\delta_l = B_l$. Indeed, in view of the inequality (2.6), we obtain that, for any function $S \in Lip_L$,

$$\max_{1 \leq l \leq n} |B_l| \leq Lh = \frac{L(b-a)}{2n}.$$

Therefore, in this case

$$\|\delta\|_n^2 \leq \frac{L^2(b-a)^3}{4n^2}. \quad (3.4)$$

Moreover from (2.14), in view of condition (2.10), we obtain that

$$\max_{1 \leq l \leq n} \sigma_l^2 \leq \frac{4}{(b-a)\epsilon_T} \leq \max\left(\frac{4}{(b-a)\epsilon_T}, 1\right) = \varsigma_n. \quad (3.5)$$

We make use of the trigonometric basis $(\phi_j)_{j \geq 1}$ in $\mathcal{L}_2[a, b]$ with

$$\phi_1 = \frac{1}{\sqrt{b-a}}, \quad \phi_j(x) = \sqrt{\frac{2}{b-a}} Tr_j(2\pi[j/2]v(x)), \quad j \geq 2, \quad (3.6)$$

where the function $Tr_j(x) = \cos(x)$ for even j and $Tr_j(x) = \sin(x)$ for odd j ; $v(x) = (x-a)/(b-a)$. This basis is orthonormal for the empiric inner product generated by the sieve (2.5), i.e. for any $1 \leq i, j \leq n$,

$$(\phi_i, \phi_j)_n = \frac{b-a}{n} \sum_{l=1}^n \phi_i(x_l) \phi_j(x_l) = \mathbf{K}r_{ij}, \quad (3.7)$$

where $\mathbf{K}r_{ij}$ is Kronecker's symbol.

By making use of this basis we apply the discrete Fourier transformation to (3.1) and we obtain the Fourier coefficients

$$\hat{\theta}_{j,n} = \frac{b-a}{n} \sum_{l=1}^n Y_l \phi_j(x_l), \quad \theta_{j,n} = \frac{b-a}{n} \sum_{l=1}^n S(x_l) \phi_j(x_l).$$

From (3.1) it follows directly that these Fourier coefficients satisfy the following equation

$$\hat{\theta}_{j,n} = \theta_{j,n} + \zeta_{j,n} \quad (3.8)$$

with

$$\zeta_{j,n} = \sqrt{\frac{b-a}{n}} \xi_{j,n} + \delta_{j,n},$$

where

$$\xi_{j,n} = \sqrt{\frac{b-a}{n}} \sum_{l=1}^n \sigma_l \xi_l \phi_j(x_l) \quad \text{and} \quad \delta_{j,n} = \frac{b-a}{n} \sum_{l=1}^n \delta_l \phi_j(x_l).$$

Notice that inequality (3.3) and the Bouniakovskii-Cauchy-Schwarz inequality imply that

$$|\delta_{j,n}| \leq \|\delta\|_n \|\phi_j\|_n = \|\delta\|_n. \quad (3.9)$$

We estimate the function S on the sieve (2.5) by the weighted least squares estimator

$$\hat{S}_\lambda(x_l) = \sum_{j=1}^n \lambda(j) \hat{\theta}_{j,n} \phi_j(x_l) \mathbf{1}_\Gamma, \quad 1 \leq l \leq n, \quad (3.10)$$

where the weight vector $\lambda = (\lambda(1), \dots, \lambda(n))'$ belongs to some finite set Λ into \mathbb{R}^n , the prime denotes the transposition. We set for any $x \in [a, b]$

$$\hat{S}_\lambda(x) = \hat{S}_\lambda(x_1) \mathbf{1}_{\{a \leq x \leq x_1\}} + \sum_{l=2}^n \hat{S}_\lambda(x_l) \mathbf{1}_{\{x_{l-1} < x \leq x_l\}}. \quad (3.11)$$

Let ν be the cardinal number of the set Λ . We suppose that the components of the weight vector $\lambda = (\lambda(1), \dots, \lambda(n))'$ are such that $0 \leq \lambda(j) \leq 1$. By setting $\lambda^2 = (\lambda^2(1), \dots, \lambda^2(n))'$ we define the following sets

$$\Lambda_1 = \{\lambda^2, \lambda \in \Lambda\} \quad \text{and} \quad \Lambda_2 = \Lambda \cup \Lambda_1. \quad (3.12)$$

We set

$$\chi_j(\lambda) = \mathbf{1}_{\{\lambda_j > 0\}}, \quad \chi_* = \max_{\lambda \in \Lambda} \sum_{j=1}^n \chi_j(\lambda) \quad \text{and} \quad \mu_n = \frac{\chi_*}{\sqrt{n}}. \quad (3.13)$$

Moreover, for the basis functions (3.6) we shall use the following upper bound

$$\lambda_* = \max_{\lambda \in \Lambda_1} \sup_{a \leq x \leq b} \left| \sum_{j=1}^n \lambda(j) \bar{\phi}_j(x) \right|, \quad (3.14)$$

where $\bar{\phi}_j = (b-a)\phi_j^2 - 1$. Now we have to write a rule to choose a weight vector $\lambda \in \Lambda$. It is obviously, that the best way is to minimize with respect to λ the empiric squared error

$$\text{Err}_n(\lambda) = \|\hat{S}_\lambda - S\|_n^2,$$

which in our case is equal to

$$\text{Err}_n(\lambda) = \sum_{j=1}^n \lambda^2(j) \hat{\theta}_{j,n}^2 - 2 \sum_{j=1}^n \lambda(j) \hat{\theta}_{j,n} \theta_{j,n} + \sum_{j=1}^n \theta_{j,n}^2. \quad (3.15)$$

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

$$\tilde{\theta}_{j,n} = \hat{\theta}_{j,n}^2 - \frac{b-a}{n} s_{j,n}$$

with

$$s_{j,n} = \frac{b-a}{n} \sum_{l=1}^n \sigma_l^2 \phi_j^2(x_l). \quad (3.16)$$

For this substitution in the empiric squared error one has to pay a penalty.

Finally, we define the cost function by the following way

$$J_n(\lambda) = \sum_{j=1}^n \lambda^2(j) \hat{\theta}_{j,n}^2 - 2 \sum_{j=1}^n \lambda(j) \tilde{\theta}_{j,n} + \rho P_n(\lambda), \quad (3.17)$$

where the penalty term we define as

$$P_n(\lambda) = \frac{(b-a)|\lambda|^2 s_n}{n}, \quad s_n = \frac{1}{n} \sum_{l=1}^n \sigma_l^2, \quad |\lambda|^2 = \sum_{j=1}^n \lambda^2(j)$$

and $0 < \rho < 1$ is some positive constant which will be chosen later. We set

$$\hat{\lambda} = \operatorname{agrmin}_{\lambda \in \Lambda} J_n(\lambda) \quad (3.18)$$

and define an estimator of S at the model (3.1) as

$$\hat{S}_*(x) = \hat{S}_{\hat{\lambda}}(x) \quad \text{for } a \leq x \leq b. \quad (3.19)$$

In the Oracle inequality we make use of the following function

$$\Upsilon_n^*(\rho) = v_1^*(\rho) \varsigma_n (1 + \mu_n) \nu + v_2^*(\rho) \varsigma_n \lambda_* + v_3^*(\rho) \chi_* \delta_n^* \quad (3.20)$$

with $\delta_n^* = n \mathbf{E}_S \mathbf{1}_\Gamma \|\delta\|_n^2$,

$$v_1^*(\rho) = \frac{16(b-a)(6+3\rho)}{\rho}, \quad v_2^*(\rho) = \frac{4\rho(b-a)}{3} \quad \text{and} \quad v_3^*(\rho) = \frac{4(6+\rho)}{\rho}.$$

Now with the help of $\Upsilon_n^*(\rho)$ we introduce the function which describes the second term in the upper bound of the Oracle inequality :

$$\Psi_n(\rho) = \psi_1(\rho) \Upsilon_n^*(\rho) + \psi_2(\rho) \varsigma_n (2\nu + \rho^2 \lambda_*) + \psi_3(\rho) \chi_* \delta_n^*, \quad (3.21)$$

where

$$\psi_1(\rho) = \frac{1-2\rho}{2\rho(1-6\rho)}, \quad \psi_2(\rho) = \frac{b-a}{2\rho(1-6\rho)} \quad \text{and} \quad \psi_3(\rho) = \frac{4+\rho}{2\rho(1-6\rho)}.$$

Theorem 3.1. For any $n \geq 3$ and $0 < \rho < 1/6$, the estimator \hat{S}_* satisfies the following Oracle inequality

$$\begin{aligned} \mathbf{E}_S \|\hat{S}_* - S\|_n^2 &\leq \frac{1+3\rho}{1-6\rho} \min_{\lambda \in \Lambda} \mathbf{E}_S \|\hat{S}_\lambda - S\|_n^2 + \frac{\Psi_n(\rho)}{n} \\ &\quad + \left(\kappa_n(\rho) + \frac{\rho}{1-6\rho} \|S\|_n^2 \right) \sqrt{\mathbf{P}_S(\Gamma^c)}, \end{aligned} \quad (3.22)$$

where

$$\kappa_n(\rho) = (\sqrt{3} + 1) \frac{\rho}{1-6\rho} \mu_n^* \quad \text{and} \quad \mu_n^* = \frac{(b-a)\varsigma_n \chi_*}{n}.$$

Remark 3.1. It is clear that to obtain the asymptotically optimal variance in the case of unknown regularity one can take, for example, the parameter ρ going to zero like

$$\rho = \rho_n = O\left(\frac{1}{\ln^\gamma n}\right) \quad (3.23)$$

for some $\gamma > 0$.

Now we consider the adaptive setting, i.e. we suppose the function S belongs to the Sobolev space

$$W_r^k = \{S \in \Sigma_L \cap \mathcal{C}^{k-1}([a, b]) : \sum_{j=0}^k \|S^{(j)}\|^2 \leq r\},$$

where the regularity exponent $k \geq 1$ and the Sobolev radius $r > 0$ are unknown.

We consider the weight family Λ introduced in [11]. For any $0 < \varepsilon < 1$ we define the set

$$\mathcal{A}_\varepsilon = \{1, \dots, k_*\} \times \{t_1, \dots, t_m\},$$

where $k_* = \lceil 1/\sqrt{\varepsilon} \rceil$, $t_i = i\varepsilon$, $m = \lceil 1/\varepsilon^2 \rceil$ and we take $\varepsilon = 1/\ln n$ for any $n \geq 3$.

For any $\alpha = (\beta, t) \in \mathcal{A}_\varepsilon$ we define the weight vector $\lambda_\alpha = (\lambda_\alpha(1), \dots, \lambda_\alpha(n))'$ with

$$\lambda_\alpha(j) = \begin{cases} 1, & \text{for } 1 \leq j \leq j_0, \\ (1 - (j/\omega_\alpha)^\beta)_+, & \text{for } j_0 < j \leq n, \end{cases} \quad (3.24)$$

where $j_0 = j_0(\alpha) = [\omega_\alpha / \ln n] + 1$,

$$\omega_\alpha = (A_\beta t n)^{1/(2\beta+1)} \quad \text{and} \quad A_\beta = \frac{(b-a)^{2\beta+1}(\beta+1)(2\beta+1)}{\pi^{2\beta}\beta}.$$

Therefore,

$$\Lambda = \{\lambda_\alpha, \alpha \in \mathcal{A}_\varepsilon\} \quad (3.25)$$

and $\nu = k_* m \leq \ln^{5/2} n$. Note that in this case

$$\sum_{j=1}^n \chi_j(\lambda) = \sum_{j=1}^n \mathbf{1}_{\{\lambda_\alpha(j) > 0\}} \leq \omega_\alpha.$$

Therefore,

$$\chi_* \leq \max_\alpha \omega_\alpha \leq (b-a)\pi^{-2/3}(6n \ln n)^{1/3}.$$

This yields

$$\lim_{n \rightarrow \infty} \mu_n = 0 \quad \text{and} \quad \lim_{n \rightarrow \infty} \mu_n^* = 0.$$

Moreover, if the sequence $(\delta_l)_{1 \leq l \leq n}$ satisfies inequality (3.4) we deduce that

$$\lim_{n \rightarrow \infty} \chi_* \delta_n^* = 0. \quad (3.26)$$

By Lemma 6.2 from [11] we get for any integer $m \geq 0$ and $a \leq x \leq b$

$$\sup_{N \geq 3} \sup_{1 \leq n < N} \left| \sum_{j=n+1}^N \left(\frac{j}{N}\right)^m \bar{\phi}_j(x) \right| \leq 2^{m+1}.$$

Therefore, for any $\alpha \in \mathcal{A}_\varepsilon$

$$\sup_{a \leq x \leq b} \left| \sum_{j=1}^n \lambda_\alpha^2(j) \bar{\phi}_j(x) \right| \leq 2 + 2 \cdot 2^{k_*+1} + 2^{2k_*+1}.$$

This means that for any $\delta > 0$

$$\lim_{n \rightarrow \infty} \frac{\lambda_*}{n^\delta} = 0.$$

Moreover, if we choose the parameter ρ as in (3.23) and if we assume as in (3.5) that the parameter ς_n is slowly increasing in n , i.e. for any $\delta > 0$

$$\lim_{n \rightarrow \infty} \frac{\varsigma_n}{n^\delta} = 0$$

we obtain that for any $\delta > 0$

$$\lim_{n \rightarrow \infty} \frac{\Psi_n(\rho)}{n^\delta} = 0. \quad (3.27)$$

Now we consider the estimation problem (1.1) via model (1.5). By denoting $\hat{S}_\alpha = \hat{S}_{\lambda_\alpha}$ We make use of estimating procedure (3.18)–(3.19) with the weight coefficients set (3.25), i.e. denoting $\hat{S}_\alpha = \hat{S}_{\lambda_\alpha}$ we set $\hat{S}_* = \hat{S}_{\hat{\alpha}}$ with

$$\hat{\alpha} = \operatorname{agrmin}_{\alpha \in \mathcal{A}_\varepsilon} J_n(\lambda_\alpha).$$

For this procedure we show the following Oracle inequality.

Theorem 3.2. *Assume that $S \in W_r^k$ with unknown parameters $r > 0$ and $k \geq 1$. Then the procedure \hat{S}_* with $n = 2[(T-1)/2] + 1$, for any $T \geq 32$ and $0 < \rho < 1/6$, satisfies the following inequality*

$$\mathcal{R}(\hat{S}_*, S) \leq \frac{(1+\rho)^2(1+3\rho)}{1-6\rho} \min_{\alpha \in \mathcal{A}_\varepsilon} \mathcal{R}(\hat{S}_\alpha, S) + \frac{\mathcal{B}_T(\rho)}{n}, \quad (3.28)$$

where

$$\mathcal{B}_T(\rho) = (1+\rho)(\Psi_n(\rho) + g_T) + \frac{(1+\rho)^2 L^2 (b-a)^3}{\rho(1-6\rho)n}$$

with

$$g_T = n(\kappa_n(\rho) + s_*(b-a))\sqrt{\Pi_T}.$$

Moreover, for any $0 < \rho < 1/6$ and any $\delta > 0$,

$$\lim_{T \rightarrow \infty} \frac{\mathcal{B}_T(\rho)}{T^\delta} = 0. \quad (3.29)$$

Remark 3.2. Note that to obtain the inequality (3.28) we make use of the method proposed in [11] and based on the nonasymptotic approach from [2] and [4].

Remark 3.3. In Part 2 of this research we will show the obtained oracle inequality is sharp in the sense that it provides the asymptotic efficiency of the estimator \hat{S}_{α^*} . Moreover, we will find the Pinsker constant and will prove that the asymptotic quadratic risk of the above estimator coincides with this constant.

4 Proofs

4.1 Proof of Theorem 3.1

First of all, note that on the set Γ we can represent the empiric squared error $\text{Err}_n(\lambda)$ by the following way

$$\text{Err}_n(\lambda) = J_n(\lambda) + 2 \sum_{j=1}^n \lambda(j) \theta'_{j,n} + \|S\|_n^2 - \rho P_n(\lambda) \quad (4.1)$$

with $\theta'_{j,n} = \tilde{\theta}_{j,n} - \theta_{j,n} \hat{\theta}_{j,n}$. From (3.8) we find that

$$\theta'_{j,n} = \theta_{j,n} \zeta_{j,n} + \frac{b-a}{n} \tilde{\xi}_{j,n} + 2 \sqrt{\frac{b-a}{n}} \xi_{j,n} \delta_{j,n} + \delta_{j,n}^2,$$

where $\tilde{\xi}_{j,n} = \xi_{j,n}^2 - s_{j,n}$. We set

$$\tilde{\Delta}(\lambda) = \frac{b-a}{n} \sum_{j=1}^n \lambda(j) \tilde{\xi}_{j,n}, \quad \Delta(\lambda) = 2 \sqrt{\frac{b-a}{n}} \sum_{j=1}^n \lambda(j) \xi_{j,n} \delta_{j,n}. \quad (4.2)$$

We remind that for any $\epsilon > 0$ and for any x, y

$$2xy \leq \epsilon x^2 + \epsilon^{-1}y^2. \quad (4.3)$$

Therefore, for some $0 < \epsilon < 1$ with help inequality (3.9) we can estimate the last term as

$$\begin{aligned} |\Delta(\lambda)| &\leq \epsilon \frac{b-a}{n} \sum_{j=1}^n \lambda^2(j) \xi_{j,n}^2 + \frac{1}{\epsilon} \sum_{j=1}^n \chi_j \delta_{j,n}^2 \\ &\leq \epsilon \frac{b-a}{n} \sum_{j=1}^n \lambda^2(j) s_{j,n} + \epsilon \tilde{\Delta}(\lambda^2) + \frac{1}{\epsilon} \chi_* \|\delta\|_n^2, \end{aligned}$$

where the vector λ^2 is defined in (3.12). Taking into account that

$$\left| \sum_{j=1}^n \lambda^2(j) s_{j,n} - |\lambda|^2 s_n \right| = \left| \frac{1}{n} \sum_{l=1}^n \sigma_l^2 \sum_{j=1}^n \lambda^2(j) \bar{\phi}_j(x_l) \right| \leq \varsigma_n \lambda_* \quad (4.4)$$

we get the following upper bound for $\Delta(\lambda)$, i.e.,

$$\begin{aligned} |\Delta(\lambda)| &\leq \epsilon P_n(\lambda) + \epsilon \tilde{\Delta}(\lambda^2) + \epsilon \frac{(b-a)\varsigma_n \lambda_*}{n} \\ &\quad + \frac{1}{\epsilon} \chi_* \|\delta\|_n^2. \end{aligned} \quad (4.5)$$

Now putting

$$M(\lambda) = \sum_{j=1}^n \lambda(j) \theta_{j,n} \zeta_{j,n} \quad \text{and} \quad \Delta_1(\lambda) = \sum_{j=1}^n \lambda(j) \delta_{j,n}^2 \quad (4.6)$$

we rewrite (4.1) by the following way

$$\begin{aligned} \text{Err}_n(\lambda) &= J_n(\lambda) + 2M(\lambda) + 2\tilde{\Delta}(\lambda) \\ &\quad + 2\Delta(\lambda) + 2\Delta_1(\lambda) + \|S\|_n^2 - \rho P_n(\lambda) \end{aligned} \quad (4.7)$$

Note now that, we can represent $\tilde{\xi}_{j,n}$ as

$$\begin{aligned} \tilde{\xi}_{j,n} &= 2(b-a) \sum_{l=2}^n u_{j,l} \xi_l + \frac{b-a}{n} \sum_{l=1}^n \sigma_l^2 \phi_j^2(x_l) \tilde{\xi}_l \\ &= 2U_{j,n} + V_{j,n}, \end{aligned} \quad (4.8)$$

where $\tilde{\xi}_l = \xi_l^2 - 1$ and

$$u_{j,l} = \frac{1}{n} \sigma_l \phi_j(x_l) \sum_{r=1}^{l-1} \sigma_r \phi_j(x_r) \xi_r.$$

To study the function $\tilde{\Delta}(\lambda)$ we set

$$\bar{U}(\lambda) = \frac{1}{\sqrt{s_n}} \sum_{j=1}^n \bar{\lambda}(j) U_{j,n} \mathbf{1}_{\{s_n > 0\}} \quad \text{and} \quad \bar{V}(\lambda) = \sum_{j=1}^n \lambda(j) V_{j,n},$$

where $\bar{\lambda}(j) = \lambda(j)/|\lambda|$. Therefore, we can represent $\tilde{\Delta}(\lambda)$ as

$$\tilde{\Delta}(\lambda) = 2\sqrt{\frac{b-a}{n}} \bar{U}(\lambda) \sqrt{P_n(\lambda)} + \frac{b-a}{n} \bar{V}(\lambda).$$

In Appendix A.2 we show that

$$\sup_{\lambda \in \mathbb{R}^n} \mathbf{E}_S \bar{U}^2(\lambda) \leq 2\varsigma_n. \quad (4.9)$$

Therefore, in view of definition (3.12) we find that

$$\mathbf{E}_S \sup_{\lambda \in \Lambda_2} \bar{U}^2(\lambda) \leq \sum_{\lambda \in \Lambda_2} \mathbf{E}_S \bar{U}^2(\lambda) \leq 4\varsigma_n \nu. \quad (4.10)$$

Moreover, taking into account that

$$\mathbf{E}_S V_{j,n}^2 \leq \frac{4\sigma_*^2}{n},$$

we get that

$$\sup_{\lambda \in \Lambda_2} \mathbf{E}_S |\bar{V}(\lambda)| \leq 2\varsigma_n \mu_n$$

and similarly to (4.10) we obtain that

$$\mathbf{E}_S \sup_{\lambda \in \Lambda_2} |\bar{V}(\lambda)| \leq 4\varsigma_n \mu_n \nu.$$

Therefore, if we set

$$\tilde{\Delta}_* = \sup_{\lambda \in \Lambda_2} \bar{U}^2(\lambda) + \sup_{\lambda \in \Lambda_2} |\bar{V}(\lambda)|,$$

then we get that

$$\mathbf{E}_S \tilde{\Delta}_* \leq 4 \varsigma_n (1 + \mu_n) \nu. \quad (4.11)$$

Moreover, taking into account that $P_n(\lambda^2) \leq P_n(\lambda)$ we obtain that for any $\lambda \in \Lambda_2$ and $0 < \epsilon < 1$

$$|\tilde{\Delta}(\lambda)| \leq \epsilon P_n(v_\lambda) + \frac{(b-a)(1+\epsilon)}{n\epsilon} \tilde{\Delta}_* \quad (4.12)$$

with $v_\lambda = \lambda$ for $\lambda \in \Lambda$ and $v_\lambda = (\sqrt{\lambda(1)}, \dots, \sqrt{\lambda(1)})'$ for $\lambda \in \Lambda_1$. We remind that the set Λ_1 is defined in (3.12). Therefore, from (4.5) replacing ϵ by ϵ^2 we can get for any $\lambda \in \Lambda$ the following upper bound for $\Delta(\lambda)$

$$\begin{aligned} |\Delta(\lambda)| &\leq 2\epsilon P_n(\lambda) + \frac{(b-a)(1+\epsilon)}{n} \tilde{\Delta}_* \\ &\quad + \epsilon \frac{(b-a)\varsigma_n \lambda_*}{n} + \frac{b-a}{\epsilon} \chi_* \|\delta\|_n^2. \end{aligned}$$

Moreover, directly from (3.9) we find that

$$\sup_{\lambda \in \Lambda} |\Delta_1(\lambda)| \leq \chi_* \|\delta\|_n^2.$$

Therefore, setting

$$\Delta_2(\lambda) = 2\tilde{\Delta}(\lambda) + 2\Delta(\lambda) + 2\Delta_1(\lambda),$$

we can estimate this term for any $\lambda \in \Lambda$ as

$$\begin{aligned} |\Delta_2(\lambda)| &\leq 6\epsilon P_n(\lambda) + \frac{2(b-a)(1+3\epsilon)}{\epsilon n} \tilde{\Delta}_* \\ &\quad + \frac{2\epsilon(b-a)}{n} \varsigma_n \lambda_* + \frac{2(1+\epsilon)}{\epsilon} \chi_* \|\delta\|_n^2. \end{aligned} \quad (4.13)$$

Now from (4.1) we obtain that for some fixed λ_0 from Λ

$$\begin{aligned} \text{Err}_n(\hat{\lambda}) - \text{Err}_n(\lambda_0) &= J(\hat{\lambda}) - J(\lambda_0) + 2M(\hat{\vartheta}) \\ &\quad + \Delta_2(\hat{\lambda}) - \rho P_n(\hat{\lambda}) - \Delta_2(\lambda_0) + \rho P_n(\lambda_0), \end{aligned}$$

where $\hat{\vartheta} = \hat{\lambda} - \lambda_0$. Therefore by definition of $\hat{\lambda}$ in (3.18) and putting $\epsilon = \rho/6$ we obtain that on the set Γ

$$\text{Err}_n(\hat{\lambda}) \leq \text{Err}_n(\lambda_0) + 2M(\hat{\vartheta}) + \Upsilon_n(\rho) + 2\rho P_n(\lambda_0), \quad (4.14)$$

where

$$\begin{aligned} \Upsilon_n(\rho) &= \frac{4(b-a)(6+3\rho)}{n\rho} \tilde{\Delta}_* + \frac{4\rho(b-a)}{3n} \varsigma_n \lambda_* \\ &+ \frac{4(6+\rho)}{\rho} \chi_* \|\delta\|_n^2. \end{aligned}$$

From (4.11) it follows that

$$\mathbf{E}_S \Upsilon_n(\rho) \leq \frac{\Upsilon_n^*(\rho)}{n}, \quad (4.15)$$

where the function $\Upsilon_n^*(\rho)$ is defined in (3.20).

Now we study the second term in the right-hand part of the inequality (4.14). Firstly for any weight vector $\lambda \in \Lambda$ we set $\vartheta = \lambda - \lambda_0$. Then we decompose this term as

$$M(\vartheta) = \sqrt{b-a} Z(\vartheta) + N(\vartheta) \quad (4.16)$$

with

$$Z(\vartheta) = n^{-1/2} \sum_{j=1}^n \vartheta(j) \theta_{j,n} \xi_{j,n} \quad \text{and} \quad N(\vartheta) = \sum_{j=1}^n \vartheta(j) \theta_{j,n} \delta_{j,n}.$$

We define now the weighted discrete Fourier transformation of S , i.e. we set

$$S_\lambda = \sum_{j=1}^n \lambda(j) \theta_{j,n} \phi_j. \quad (4.17)$$

It is easy to see that in this case

$$\mathbf{E}_S Z^2(\vartheta) = \frac{b-a}{n^2} \sum_{l=1}^n \mathbf{E}_S \sigma_l^2 \left(\sum_{j=1}^n \vartheta(j) \theta_{j,n} \phi_j(x_l) \right)^2 \leq \varsigma_n d(\vartheta), \quad (4.18)$$

where

$$d(\vartheta) = \frac{\|S_\vartheta\|_n^2}{n}.$$

Moreover, note that we can represent $\vartheta(j) = \lambda(j) - \lambda_0(j)$ as

$$\vartheta(j) = \vartheta(j)\chi_j^* \quad \text{with} \quad \chi_j^* = \max(\mathbf{1}_{\{\lambda(j)>0\}}, \mathbf{1}_{\{\lambda_0(j)>0\}}).$$

Therefore, by inequality (4.3) with $\epsilon = \rho$, we can estimate $N(\vartheta)$ as follows

$$\begin{aligned} 2N(\vartheta) &= 2 \sum_{j=1}^n \vartheta(j) \theta_{j,n} \chi_j^* \delta_{j,n} \\ &\leq \rho \|S_\vartheta\|_n^2 + \frac{2\chi_* \|\delta\|_n^2}{\rho}. \end{aligned} \quad (4.19)$$

Setting

$$Z^* = \sup_{\vartheta \in \Theta} \frac{Z^2(\vartheta)}{d(\vartheta)} \quad \text{with} \quad \Theta = \Lambda - \lambda_0,$$

we obtain on the set Γ

$$2M(\vartheta) \leq 2\rho \|S_\vartheta\|_n^2 + \frac{(b-a)Z^*}{n\rho} + \frac{2\chi_* \|\delta\|_n^2}{n}. \quad (4.20)$$

Note now that from (4.18) it follows that

$$\mathbf{E}_S Z^* \leq \nu \zeta_n. \quad (4.21)$$

Now we estimate the first term in the right-hand part of the inequality (4.20).

On the set Γ we have

$$\begin{aligned} \|S_\vartheta\|_n^2 - \|\hat{S}_\vartheta\|_n^2 &= \sum_{j=1}^n \vartheta^2(j) (\theta_{j,n}^2 - \hat{\theta}_{j,n}^2) \leq -2 \sum_{j=1}^n \vartheta^2(j) \theta_{j,n} \zeta_{j,n} \\ &= -2\sqrt{b-a} Z_1(\vartheta) - 2N_1(\vartheta), \end{aligned} \quad (4.22)$$

where

$$Z_1(\vartheta) = \frac{1}{\sqrt{n}} \sum_{j=1}^n \vartheta^2(j) \theta_{j,n} \xi_{j,n} \quad \text{and} \quad N_1(\vartheta) = \sum_{j=1}^n \vartheta^2(j) \theta_{j,n} \delta_{j,n}.$$

By the same way, taking into account that $|\vartheta(j)| \leq 1$, we find that on the set Γ

$$\mathbf{E}_S Z_1^2(\vartheta) \leq \varsigma_n d(\vartheta).$$

Similarly, we set

$$Z_1^* = \sup_{\vartheta \in \Theta} \frac{Z_1^2(\vartheta)}{d(\vartheta)}$$

and we obtain that on the set Γ

$$\mathbf{E}_S Z_1^* \leq \nu \varsigma_n. \quad (4.23)$$

Similarly to (4.17) we estimate the second term in (4.22) as

$$2N_1(\vartheta) \leq \rho \|S_\vartheta\|_n^2 + \frac{2\chi_* \|\delta\|_n^2}{\rho}.$$

Therefore, on the set Γ

$$\|S_\vartheta\|_n^2 \leq \|\hat{S}_\vartheta\|_n^2 + 2\rho \|S_\vartheta\|_n^2 + \frac{(b-a)Z_1^*}{n\rho} + \frac{2\chi_* \|\delta\|_n^2}{\rho},$$

i.e.

$$\|S_\vartheta\|_n^2 \leq \frac{1}{1-2\rho} \|\hat{S}_\vartheta\|_n^2 + \frac{1}{(1-2\rho)\rho} \left(\frac{(b-a)Z_1^*}{n} + 2\chi_* \|\delta\|_n^2 \right). \quad (4.24)$$

By applying this inequality in (4.20) and by putting $Z_2^* = Z^* + Z_1^*$ we obtain that on the set Γ

$$\begin{aligned} 2M(\vartheta) &\leq \frac{2\rho}{1-2\rho} \|\hat{S}_\vartheta\|_n^2 + \frac{1}{\rho(1-2\rho)} \left(\frac{(b-a)Z_2^*}{n} + 2\chi_* \|\delta\|_n^2 \right) \\ &\leq \frac{4\rho(\text{Err}_n(\lambda) + \text{Err}_n(\lambda_0))}{1-2\rho} + \frac{1}{\rho(1-2\rho)} \left(\frac{(b-a)Z_2^*}{n} + 2\chi_* \|\delta\|_n^2 \right). \end{aligned}$$

Therefore (4.14) implies that

$$\begin{aligned} \text{Err}_n(\hat{\lambda}) \mathbf{1}_\Gamma &\leq \frac{1+2\rho}{1-6\rho} \text{Err}_n(\lambda_0) \mathbf{1}_\Gamma + \frac{1-2\rho}{1-6\rho} \Upsilon_n(\rho) \mathbf{1}_\Gamma \\ &\quad + \frac{1}{\rho(1-6\rho)} \left(\frac{(b-a)Z_2^*}{n} + 2\chi_* \|\delta\|_n^2 \right) \mathbf{1}_\Gamma + \frac{2\rho(1-2\rho)}{1-6\rho} P_n(\lambda_0) \mathbf{1}_\Gamma, \end{aligned}$$

Now by inequalities (4.21)–(4.23) we get that

$$\begin{aligned} \mathbf{E}_S \text{Err}_n(\hat{\lambda}) \mathbf{1}_\Gamma &\leq \frac{1+2\rho}{1-6\rho} \mathbf{E}_S \text{Err}_n(\lambda_0) \mathbf{1}_\Gamma + \frac{1-2\rho}{n(1-6\rho)} \Upsilon_n^*(\rho) \\ &\quad + \frac{2\nu\varsigma_n(b-a) + 2\chi_*\delta_n^*}{n\rho(1-6\rho)} + \frac{\rho(1-2\rho)}{1-6\rho} \mathbf{E}_S \mathbf{1}_\Gamma P_n(\lambda_0). \end{aligned}$$

By applying here Lemma A.1 with $\epsilon = 2\rho$ we get that

$$\mathbf{E}_S \text{Err}_n(\hat{\lambda}) \mathbf{1}_\Gamma \leq \frac{1+3\rho}{1-6\rho} \mathbf{E}_S \text{Err}_n(\lambda_0) \mathbf{1}_\Gamma + \frac{\Psi_n(\rho)}{n} + \kappa_n(\rho) \sqrt{\mathbf{P}_S(\Gamma^c)},$$

where the functions $\Psi_n(\cdot)$ and $\kappa_n(\cdot)$ are defined in (3.21)–(3.22). Replacing here

$$\mathbf{E}_S \text{Err}_n(\hat{\lambda}) \mathbf{1}_\Gamma \quad \text{and} \quad \mathbf{E}_S \text{Err}_n(\lambda_0) \mathbf{1}_\Gamma$$

by

$$\mathbf{E}_S \|\hat{S}_* - S\|_n^2 - \|S\|_n^2 \mathbf{P}_S(\Gamma^c) \quad \text{and} \quad \mathbf{E}_S \|\hat{S}_{\lambda_0} - S\|_n^2 - \|S\|_n^2 \mathbf{P}_S(\Gamma^c),$$

respectively, we come to the inequality (3.22). \square

4.2 Proof of Theorem 3.2

We apply now Theorem 3.1 to the estimation problem at the model (1.1) via model (1.5) with the estimation family (3.25). First, note that, for any function $S \in \Sigma_L$ and for any pointwise estimator $\hat{S}(\cdot)$ of the form (3.11) it is easy to see that

$$\|\hat{S} - S\|^2 \leq (1+\rho) \|\hat{S} - S\|_n^2 + (1+\rho^{-1}) L^2 (b-a)^3 \frac{1}{n^2}$$

and

$$\|\hat{S} - S\|_n^2 \leq (1+\rho) \|\hat{S} - S\|^2 + (1+\rho^{-1}) L^2 (b-a)^3 \frac{1}{n^2}.$$

Moreover, inequality (2.4) implies

$$\|S\|_n^2 \leq s_*(b-a)$$

for any S from Σ_L . Therefore inequality (3.22) implies directly (3.28). Hence Theorem 3.2. \square

A Appendix

A.1 Proof of Proposition 2.1

First note that for this proposition we make use the concentration inequality from [13]. We remind it. For any $c > 0$ and for any function $\psi : \mathbb{R} \rightarrow \mathbb{R}$ such that

$$\sup_{y \in \mathbb{R}} |\psi(y)| < \infty \quad \text{and} \quad \int_{-\infty}^{+\infty} |\psi(y)| dy \leq c < \infty \quad (\text{A.1})$$

we set

$$\mathbf{M}_S(\psi) = \int_{-\infty}^{+\infty} \psi(y) q_S(y) dy$$

and

$$\Delta_{t_0, T}(\psi) = \frac{1}{\sqrt{T}} \int_{t_0}^T (\psi(y_t) - \mathbf{M}_S(\psi)) dt.$$

We will denote by $\Delta_t(\psi) = \Delta_{0, t}(\psi)$. Note that now by Theorem 3.2 from [13] for any function $\psi(\cdot)$ satisfying (A.1) we have

$$\sup_{T \geq 1} \sup_{S \in \Sigma_L} \mathbf{E}_S e^{\gamma \Delta_T^2(\psi)} \leq 4, \quad (\text{A.2})$$

where the constant $\gamma = \gamma(c, L)$ is defined by (2.15). Therefore taking into account the inequality

$$|\Delta_{t_0, T}(\psi)| \leq |\Delta_T(\psi)| + |\Delta_{t_0}(\psi)|,$$

we obtain that for any $\mu > 0$ and for any ψ satisfying (A.1)

$$\sup_{1 \leq t_0 \leq T} \sup_{S \in \Sigma_L} \mathbf{P}_S(|\Delta_{t_0, T}(\psi)| \geq 2\mu) \leq 8 e^{-\gamma\mu^2}. \quad (\text{A.3})$$

Now we estimate the probability of Γ^c as

$$\mathbf{P}_S(\Gamma^c) \leq \sum_{k=1}^n \mathbf{P}_S(\tau_k > T).$$

By the definition of τ_k in (2.6) one has

$$\mathbf{P}_S(\tau_k > T) = \mathbf{P}_S\left(\int_{t_0}^T \psi_{h,k}(y_t) dt < (T - t_0)(2\tilde{q}_T(x_k) - \epsilon_T^2)\right),$$

where

$$\psi_{h,k}(y) = \frac{1}{h} Q\left(\frac{y - x_k}{h}\right).$$

It is easy to see that this function satisfies condition (A.1) with $c = 2$.

Moreover, notice now that we can represent the integral $\int_{t_0}^T \psi_{h,k}(y_t) dt$ as follows :

$$\int_{t_0}^T \psi_{h,k}(y_t) dt = (T - t_0) \mathbf{M}_S(\psi_{h,k}) + \sqrt{T} \Delta_{t_0, T}(\psi_{h,k}),$$

where

$$\mathbf{M}_S(\psi_{h,k}) = \int_{-\infty}^{+\infty} \psi_{h,k}(y) q_S(y) dy = \int_{-1}^1 q_S(x_k + hz) dz.$$

Taking into account that

$$2\hat{q}_T(x_k) - \mathbf{M}_S(\psi_{h,k}) = \frac{1}{\sqrt{t_0}} \Delta_{t_0}(\psi_{h,k}) \quad (\text{A.4})$$

we obtain

$$\begin{aligned} \mathbf{P}_S(\tau_k > T) &\leq \mathbf{P}_S(\hat{q}(x_k) < \epsilon_T) \\ &+ \mathbf{P}_S\left(\Delta_{t_0, T}(\psi_{h,k}) < \frac{T - t_0}{\sqrt{T}} \left(\frac{1}{\sqrt{t_0}} \Delta_{t_0}(\psi_{h,k}) - \epsilon_T^2\right)\right). \end{aligned}$$

Moreover, thanks to the conditions (2.10), the last probability can be estimated as follows :

$$\mathbf{P}_S(|\Delta_{t_0}(\psi_{h,k})| \geq t_0^{1/4}) + \mathbf{P}_S\left(|\Delta_{t_0,T}(\psi_{h,k})| \geq t_0^{1/4}/\sqrt{2}\right).$$

Finally we obtain that

$$\begin{aligned} \mathbf{P}_S(\tau_k > T) &\leq \mathbf{P}_S(\hat{q}(x_k) < \epsilon_T) + \mathbf{P}_S(|\Delta_{t_0}(\psi_{h,k})| \geq t_0^{1/4}) \\ &\quad + \mathbf{P}_S\left(|\Delta_{t_0,T}(\psi_{h,k})| \geq t_0^{1/4}/\sqrt{2}\right). \end{aligned} \quad (\text{A.5})$$

Let us estimate now the first probability in the left hand of this inequality.

Taking into account that $\mathbf{M}_S(\psi_{h,k}) \geq 2q_*$ we get

$$\begin{aligned} \mathbf{P}_S(\hat{q}(x_k) < \epsilon_T) &= \mathbf{P}_S\left(\frac{1}{t_0} \int_0^{t_0} \psi_{h,k}(y_t) dt < 2\epsilon_T\right) \\ &= \mathbf{P}_S\left(\Delta_{t_0}(\psi_{h,k}) < (2\epsilon_T - \mathbf{M}_S(\psi_{h,k}))\sqrt{t_0}\right) \\ &\leq \mathbf{P}_S\left(\Delta_{t_0}(\psi_{h,k}) < 2(\epsilon_T - q_*)\sqrt{t_0}\right) \end{aligned}$$

Therefore if $\epsilon_T \leq q_*/2$ then the inequality (A.2) implies

$$\mathbf{P}_S(\hat{q}(x_k) < \epsilon_T) \leq \mathbf{P}_S(|\Delta_{t_0}(\psi_{h,k})| > q_*\sqrt{t_0}) \leq 4e^{-\gamma q_*^2 t_0}. \quad (\text{A.6})$$

Now, by applying in (A.5) the inequalities (A.3) and (A.6) we obtain that for $\epsilon_T \leq q_*/2$

$$\mathbf{P}_S(\tau_k > T) \leq 16e^{-\gamma_1 \sqrt{t_0}},$$

where γ_1 is defined in (2.16). Therefore taking into account that $n \leq T$ we obtain the following upper bound

$$\mathbf{P}_S(\Gamma^c) \leq \sum_{k=1}^n \mathbf{P}_S(\tau_k > T) \leq T \mathbf{1}_{\{\epsilon_T \geq q_*/2\}} + 24T e^{-2\gamma_1 \sqrt{t_0}} \mathbf{1}_{\{\epsilon_T < q_*/2\}}.$$

Hence the condition (2.11) implies Proposition 2.1. \square

A.2 Proof of Inequality (4.9)

By putting $\alpha_l = \sum_{j=1}^n \bar{\lambda}(j) u_{j,l}$ and taking into account that the random variables $(\xi_k)_{1 \leq k \leq n}$ are independent of $(\sigma_k)_{1 \leq k \leq n}$ we obtain that

$$\mathbf{E}_S (N_1^2(\lambda) | \sigma_k, 1 \leq k \leq n) = \mathbf{1}_{\{s_n > 0\}} \frac{(b-a)^2}{s_n} \sum_{l=1}^n \hat{\alpha}_l, \quad (\text{A.7})$$

where

$$\hat{\alpha}_l = \mathbf{E}(\alpha_l^2 | \sigma_k, 1 \leq k \leq n) = \frac{\sigma_l^2}{n^2} \sum_{r=1}^{l-1} \sigma_r^2 \left(\sum_{j=1}^n \bar{\lambda}(j) \phi_j(x_l) \phi_j(x_r) \right)^2.$$

Therefore the orthonormality of the functions (ϕ_j) implies that

$$\begin{aligned} \hat{\alpha}_l &\leq \varsigma_n \frac{\sigma_l^2}{n^2} \sum_{r=1}^n \left(\sum_{j=1}^n \bar{\lambda}(j) \phi_j(x_l) \phi_j(x_r) \right)^2 \\ &= \varsigma_n \frac{\sigma_l^2}{n(b-a)} \sum_{j=1}^n \bar{\lambda}^2(j) \phi_j^2(x_l) \leq \frac{2\varsigma_n}{n(b-a)^2} \sigma_l^2. \end{aligned}$$

Now by making use of this inequality in (A.7) we get (4.9). \square

A.3 Technical lemma

Lemma A.1. *For any $n \geq 1$ and $0 < \epsilon < 1$*

$$\begin{aligned} \mathbf{E}_S P_n(\lambda) \mathbf{1}_\Gamma &\leq \frac{1}{1-\epsilon} \mathbf{E}_S \text{Err}_n(\lambda) \mathbf{1}_\Gamma + \frac{\epsilon(b-a)\varsigma_n \lambda_* + \chi_* \delta_n^*}{(1-\epsilon)\epsilon n} \\ &\quad + \frac{\sqrt{3}\mu_n^* + 2\|S\|_n \sqrt{\mu_n^*}}{1-\epsilon} \sqrt{\mathbf{P}_S(\Gamma^c)}, \end{aligned}$$

where μ_n^* is defined in (3.22).

Proof. Indeed, by the definition of $\text{Err}_n(\lambda)$ on the set Γ we have

$$\begin{aligned} \text{Err}_n(\lambda) &= \sum_{j=1}^n ((\lambda(j) - 1)\theta_{j,n} + \lambda(j)\zeta_{j,n})^2 \\ &= \sum_{j=1}^n \left((\lambda(j) - 1)\theta_{j,n} + \lambda(j)\delta_{j,n} + \lambda(j)\sqrt{\frac{b-a}{n}}\xi_{j,n} \right)^2. \end{aligned}$$

Therefore, putting

$$I_1 = \sum_{j=1}^n \lambda(j)(\lambda(j) - 1)\theta_{j,n}\xi_{j,n} \quad \text{and} \quad I_2 = \sum_{j=1}^n \lambda^2(j)\delta_{j,n}\xi_{j,n}$$

we get on the set Γ the following lower bound for the empirical risk

$$\begin{aligned} \text{Err}_n(\lambda) &\geq \frac{b-a}{n} \sum_{j=1}^n \lambda^2(j) \xi_{j,n}^2 \\ &\quad + 2\sqrt{\frac{b-a}{n}} I_1 + 2\sqrt{\frac{b-a}{n}} I_2. \end{aligned} \quad (\text{A.8})$$

Notice that the random variable $\xi_{j,n} \sim \mathcal{N}(0, s_{j,n})$ (conditionally with respect to $\sigma_1, \dots, \sigma_n$) with

$$s_{j,n} = \frac{b-a}{n} \sum_{l=1}^n \sigma_l^2 \phi_j^2(x_l) \leq \varsigma_n.$$

Thus

$$\mathbf{E}_S \xi_{j,n}^4 \leq 3\sigma_*^2. \quad (\text{A.9})$$

Let us consider the first term in (A.8). We have

$$\mathbf{E}_S \mathbf{1}_\Gamma \sum_{j=1}^n \lambda^2(j) \xi_{j,n}^2 = \mathbf{E}_S \sum_{j=1}^n \lambda^2(j) s_{j,n} - \mathbf{E}_S \mathbf{1}_{\Gamma^c} \sum_{j=1}^n \lambda^2(j) \xi_{j,n}^2.$$

Therefore, by inequalities (4.4) and (A.9) we get that

$$\mathbf{E}_S \mathbf{1}_\Gamma \sum_{j=1}^n \lambda^2(j) \xi_{j,n}^2 \geq |\lambda|^2 \mathbf{E}_S s_n \mathbf{1}_\Gamma - \varsigma_n \lambda_* - \sqrt{3} \varsigma_n \chi_* \sqrt{\mathbf{P}_S(\Gamma^c)}. \quad (\text{A.10})$$

Moreover, taking into account that $\mathbf{E}_S I_1 = 0$ we estimate $\mathbf{E}_S I_1 \mathbf{1}_\Gamma$ as

$$\begin{aligned} |\mathbf{E}_S I_1 \mathbf{1}_\Gamma| &= |\mathbf{E}_S I_1 \mathbf{1}_{\Gamma^c}| \\ &\leq \|S\|_n \left(\sum_{j=1}^n \lambda^2(j) \mathbf{E}_S s_{j,n} \right)^{1/2} \sqrt{\mathbf{P}_S(\Gamma^c)} \\ &\leq \|S\|_n \sqrt{\varsigma_n \chi_*} \sqrt{\mathbf{P}_S(\Gamma^c)}. \end{aligned} \quad (\text{A.11})$$

Let us estimate the last term in (A.8). For this we make use of inequality (4.3), i.e. for $0 < \epsilon < 1$

$$2\sqrt{\frac{b-a}{n}}|I_2| \leq \epsilon \frac{b-a}{n} \sum_{j=1}^n \lambda^2(j) \xi_{j,n}^2 + \frac{1}{n\epsilon} \chi_* \delta_n^*.$$

Therefore from (A.8) we get that

$$\text{Err}_n(\lambda) \geq (1-\epsilon) \frac{b-a}{n} \sum_{j=1}^n \lambda^2(j) \xi_{j,n}^2 + 2\sqrt{\frac{b-a}{n}} I_1 - \frac{\chi_* \delta_n^*}{n\epsilon}.$$

□

References

- [1] Arato, Kolmogorov, A.N. and Sinai, Ya. G. (1962) On parameter estimation of a complex stationary gaussian process. *Doklady Acad. SSSR* **146** 747-750.
- [2] Baron, A., Birgé, L. and Massart, P. (1999) Risk bounds for model selection via penalisation. *Probab. Theory Related Fields*, **113**, 301-413.
- [3] Birgé, L. and Massart, P. (2001) Gaussian model selection. *Journal of the European Mathematical Society*, n.3, 203-268.
- [4] Fourdrinier, D. and Pergamenshchikov, S (2007) Improved selection model method for the regression with dependent noise. - *Annals of the Institute of Statistical Mathematics*, **59** (3), p. 435-464.
- [5] Galtchouk L. and Konev, V.V. (1997) On Sequential Estimation of Parameters in Continuous-Time Stochastic Regression. *In: Statistics and Control of Stochastic Processes. The Liptser Festschrift.*, 123-138.

- [6] Galtchouk, L.I. and Konev, V.V. (2001) On Sequential Estimation of Parameters in Semimartingale Regression Models with Continuous Time Parameter. *The Annals of Statistics*, **29**, 1508-2035.
- [7] Galtchouk, L. and Pergamenshchikov, S. (2000) Estimateurs séquentiels à noyaux et aux polynômes locaux pour le problème d'estimation nonparamétrique de la dérive d'un processus de diffusion. *Prépublication de l'Institut de Recherche Mathématique Avancée de l'Université de Strasbourg*, No 2000/58.
- [8] 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.
- [9] Galtchouk, L. and Pergamenshchikov, S. (2004) Nonparametric sequential estimation of the drift in diffusion via model selection. *Mathematical Methods of Statistics*, **13**, 1, 25-49.
- [10] Galtchouk, L. and Pergamenshchikov, S. (2005) Nonparametric sequential minimax estimation of the drift in diffusion processes. *Sequential Analysis*, **24**, 3, 303-330.
- [11] Galtchouk, L. and Pergamenshchikov, S. (2005) Efficient adaptive nonparametric estimation in heteroscedastic regression models. Preprint of the Strasbourg Louis Pasteur University, IRMA, 2005/020 available at <http://www.univ-rouen.fr/LMRS/Persopage/Pergamenchtchikov>

- [12] Galtchouk, L. and Pergamenshchikov, S. (2006) Asymptotic efficient sequential kernel estimates of the drift coefficient in ergodic diffusion processes. *Statistical Inference for Stochastic Processes*, **9**, 1-16.
- [13] Galtchouk, L. and Pergamenshchikov, S. (2007) Uniform concentration inequality for ergodic diffusion processes. *Stochastic processes and their applications*, v. 7, p. 830-839.
- [14] Gihkman, I.I. and Skorohod, A.V. (1968) Stochastic differential equations. *Naukova Dumka*, Kiev.
- [15] Ibragimov, I.A. and Hasminskii, R.Z. (1979) *Statistical Estimation: Asymptotic Theory*. Springer, Berlin, New York.
- [16] Konev, V.V. and Pergamenshchikov, S.M. (1981) Sequential Estimation of the Parameters of Random Processes with Continuous Time. *Math. Statistics and Appl.*, Tomsk State University, Russia, **8**, pp. 93-101.
- [17] Konev, V.V. and Pergamenshchikov, S.M. (1992) On Truncated Sequential Estimation of the Parameters of Diffusion Processes. *Methods of Economical Analysis, Central Economical and Mathematical Institute of Russian Academy of Science, Moscow*, 3-31.
- [18] Liptser, R.Sh. and Shiryaev, A.N. (1978) *Statistics of a random process*. Springer-Verlag, Berlin, Heidelberg, New York, v.I, II.
- [19] Massart, P. (2005) A Non-asymptotic theory for model selection. *4 ECM Stockholm*, European Mathematical Society.

- [20] Novikov, A.A. (1971) Sequential estimation of the parameters of the diffusion processes. *Theory Probability and Appl.*, **16**, p. 394-396.

L.Galtchouk

Department of Mathematics

Strasbourg University

7, str. Rene Descartes

67084, Strasbourg

Cedex, France

e-mail: galtchou@math.u-strasbg.fr

S. Pergamenshchikov

Laboratoire de Mathématiques,

Avenue de l'Université, BP. 12,

Université de Rouen,

F76801, Saint Etienne du Rouvray

Cedex, France

Serge.Pergamenchtchikov@univ-rouen.fr