

Lepski, Oleg V.

**Working Paper**

## How to improve accuracy of estimation

SFB 373 Discussion Paper, No. 1997,21

**Provided in Cooperation with:**

Collaborative Research Center 373: Quantification and Simulation of Economic Processes,  
Humboldt University Berlin

*Suggested Citation:* Lepski, Oleg V. (1997) : How to improve accuracy of estimation, SFB 373 Discussion Paper, No. 1997,21, Humboldt University of Berlin, Interdisciplinary Research Project 373: Quantification and Simulation of Economic Processes, Berlin, <https://nbn-resolving.de/urn:nbn:de:kobv:11-10064073>

This Version is available at:

<https://hdl.handle.net/10419/66236>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

# HOW TO IMPROVE ACCURACY OF ESTIMATION.

LEPSKI O.V.

Humboldt Universität zu Berlin.  
Spandauer str. 1, 10178 Berlin, Germany

**ABSTRACT.** The new approach, allowed to take into account some additional information, coming from datas, is proposed. The main idea is to obtain from datas some information about structure of the model in order to improve accuracy of estimation. It seems to be important, since standard nonparametric accuracy of estimation is usually very low. To improve one statisticians often impose some additional structure on considerable model, that can lead to inadequate model. To avoid both these disadvantages special form of estimation procedure, based on some combination of adaptive technique and hypothesis testing, is applied. From mathematical point of view it leads to the consideration of new kind of minimax risks. From practical point of view it allows to improve accuracy of estimation procedures even for the cases when guess on special structure of a model turns out to be wrong.

## 1. Introduction

The paper deals with the approach, allowing to improve the quality of estimation procedures. This approach is general and one can be applied to any statistical model and to an arbitrary structural assumptions. By this reason, it is convenient to present all general definitions and to explain approach itself in terms of abstract statistical model, in other words, in terms of the sequence of statistical experiments, see Ibragimov and Khasminskii (1981). Let  $(\mathcal{X}^\varepsilon, \mathcal{B}^\varepsilon, \mathbf{P}_f^\varepsilon, f(\cdot) \in \Sigma)$  be the statistical experiment, generated by the observation  $X^\varepsilon$ . Here  $(\mathcal{X}^\varepsilon, \mathcal{B}^\varepsilon)$  be some measurable space,  $\mathbf{P}_f^\varepsilon$  is the probability measure, defined on this space,  $\varepsilon \in (0, 1)$  is the small parameter and later on the asymptotics will be studied w.r.t.  $\varepsilon \rightarrow 0$ . The set  $\Sigma$  is some given set of functions, determined on the Euclidean space  $\mathbf{R}^s$ ,  $f(\cdot) : \mathbf{R}^s \rightarrow \mathbf{R}^1$ . Here and later we assume that

$$\Sigma \subset \Sigma_s(L) = \left\{ f(\cdot) : \sup_{t \in [0,1]^s} |f(t)| \leq L < \infty \right\},$$

for some  $L > 0$ . For  $\forall 1 \leq p \leq \infty$  denote

$$\|f(\cdot)\|_p = \begin{cases} \left( \int_{[0,1]^s} |f(t)|^p dt \right)^{\frac{1}{p}} & 1 \leq p < \infty, \\ \sup_{t \in [0,1]^s} |f(t)| & p = \infty, \end{cases}$$

and consider the minimax risk on the set  $\Sigma$ :

---

The paper was printed using funds made available by the Deutsche Forschungsgemeinschaft.

$$R_\varepsilon \left( \tilde{f}_\varepsilon, \Sigma, \varphi_\varepsilon(\Sigma) \right) = \sup_{f(\cdot) \in \Sigma} \mathbf{E}_f^\varepsilon \left\{ \varphi_\varepsilon^{-1}(\Sigma) \| \tilde{f}_\varepsilon(\cdot) - f(\cdot) \|_p \right\}^q, \quad (1.1)$$

where  $\tilde{f}_\varepsilon(t) = \tilde{f}_\varepsilon(t, X^\varepsilon)$ ,  $t \in [0, 1]^s$  is some estimator, i.e. a function, defined on  $[0, 1]^s$  and being measurable w.r.t. observation  $X^\varepsilon$ ;  $\mathbf{E}_f^\varepsilon$  is mathematical expectation w.r.t. measure  $\mathbf{P}_f^\varepsilon$ ;  $q > 0$  is some fixed number, and  $\varphi_\varepsilon(\Sigma) > 0$  is normalizing factor (n.f.), i.e.  $\varphi_\varepsilon(\Sigma) \rightarrow 0$ , as  $\varepsilon \rightarrow 0$ .

Normalizing factor  $\varphi_\varepsilon(\Sigma)$  is called minimax rate of convergence (MRC) if

- $\liminf_{\varepsilon \rightarrow 0} \inf_{\tilde{f}_\varepsilon} R \left( \tilde{f}_\varepsilon, \Sigma, \varphi_\varepsilon(\Sigma) \right) > 0$ ,
- $\limsup_{\varepsilon \rightarrow 0} R \left( \hat{f}_\varepsilon, \Sigma, \varphi_\varepsilon(\Sigma) \right) < \infty$

for some estimator  $\hat{f}_\varepsilon$ , which is called asymptotically efficient (a.e.) in minimax sense.

Along the paper, the set  $\Sigma$  is assumed to be such that MRC  $\varphi_\varepsilon(\Sigma)$  and a.e.  $\hat{f}_\varepsilon(\cdot)$  exist and they are known. Note, that MRC can be treated as an accuracy of estimation. As it follows from the definition of MRC, this accuracy is attainable and unimprovable in minimax sense. However, what should a statistician do in the situation, when the accuracy of estimation is bad, say,  $\varphi_\varepsilon(\Sigma)$  tends to zero too slowly? Can one be improved? The answer on last question is, of course, negative under consideration of minimax risks of type (1.1). On the other hand, the consideration of other types of minimax risks would possibly lead to the positive solution of this problem. Before than to answer this question, we should answer another one: why do we hope that such improvement is possible on principle? One of the answers, which seems to be reasonable, consists in the following. Let us suppose we have strong suspicion (hypothesis)

$$\mathbf{H}_0 : f(\cdot) \in \Sigma_0, \Sigma_0 \subset \Sigma.$$

It is supposed to be known that there exist an estimator  $\hat{f}_\varepsilon^{(0)}$  and the n.f.  $\varphi_\varepsilon(\Sigma_0)$ , being MRC on  $\Sigma_0$ , such that

- $\limsup_{\varepsilon \rightarrow 0} R_\varepsilon \left( \hat{f}_\varepsilon^{(0)}, \Sigma_0, \varphi_\varepsilon(\Sigma_0) \right) < \infty$ ,
- $\frac{\varphi_\varepsilon(\Sigma_0)}{\varphi_\varepsilon(\Sigma)} \rightarrow 0$ , as  $\varepsilon \rightarrow 0$ ,

where for an arbitrary estimator  $\tilde{f}_\varepsilon$

$$R_\varepsilon \left( \tilde{f}_\varepsilon, \Sigma_0, \varphi_\varepsilon(\Sigma_0) \right) = \sup_{f(\cdot) \in \Sigma_0} \mathbf{E}_f^\varepsilon \left\{ \varphi_\varepsilon^{-1}(\Sigma_0) \| \tilde{f}_\varepsilon(\cdot) - f(\cdot) \|_p \right\}^q. \quad (1.2)$$

Thus, the hope on improvement of accuracy of estimation is based on the hypothesis on belonging an estimated function to the set  $\Sigma_0$ , where more precise estimation procedures are available. In this context there are, at least, two possibilities to give the mathematical sense to the words "improvement of accuracy of estimation". The first one is to use so-called adaptive approach. This approach is very popular last time and there exist a lot of publications on this topic. One could mention the papers [1]-[12], [18]-[24], [27] and [31] among others.

**1.1. Adaptive approach.** The setup of adaptive estimation, reduced to our purposes, consists in the following: one needs to find an estimator  $f_\varepsilon^{(a)}(\cdot)$  such that

$$\begin{aligned} \limsup_{\varepsilon \rightarrow 0} R_\varepsilon \left( f_\varepsilon^{(a)}, \Sigma, \varphi_\varepsilon(\Sigma) \right) &< \infty, \\ \limsup_{\varepsilon \rightarrow 0} R_\varepsilon \left( f_\varepsilon^{(a)}, \Sigma_0, \varphi_\varepsilon(\Sigma_0) \right) &< \infty. \end{aligned} \quad (1.3)$$

Any estimator, satisfying (1.3), is called adaptive. The definition of the adaptive estimator can be done in the following, equivalent (1.3), form. Put

$$\hat{\varrho}_\varepsilon(f) = \begin{cases} \varphi_\varepsilon(\Sigma_0) & f(\cdot) \in \Sigma_0, \\ \varphi_\varepsilon(\Sigma) & f(\cdot) \in \Sigma \setminus \Sigma_0, \end{cases}$$

and consider the risk, which could be called adaptive.

$$R_\varepsilon^{(a)} \left( \tilde{f}_\varepsilon, \Sigma \right) = \sup_{f(\cdot) \in \Sigma} \mathbf{E}_f^\varepsilon \left\{ \hat{\varrho}^{-1}(f) \|\tilde{f}_\varepsilon(\cdot) - f(\cdot)\|_p \right\}^q. \quad (1.4)$$

Then adaptive estimator  $f_\varepsilon^{(a)}(\cdot)$  is an estimator, providing the finiteness of the risk (1.4), i.e.

$$\limsup_{\varepsilon \rightarrow 0} R_\varepsilon^{(a)} \left( f_\varepsilon^{(a)}, \Sigma \right) < \infty, \quad (1.5)$$

see Lepski and Spokoiny (1995a) for more details on risks of type (1.4). Obviously, conditions (1.3) and (1.5) are equivalent. Note also two important facts.

First, considering risks of type (1.4), we leave the frameworks of standard minimax approach, because now n.f. can depend on function to be estimated. However, any estimator  $f_\varepsilon^{(a)}$ , satisfying (1.5), is a.e. estimator on the set  $\Sigma$  w.r.t. the minimax risk (1.1). It follows from the facts that  $\varphi_\varepsilon(\Sigma)$  is MRC on  $\Sigma$ , and  $\hat{\varrho}_\varepsilon(f) \leq \varphi_\varepsilon(\Sigma)$  for  $\forall f(\cdot) \in \Sigma$ .

Next, consider the set of pairs  $(a_\varepsilon, b_\varepsilon)$ , where  $0 < a_\varepsilon < b_\varepsilon \leq \varphi_\varepsilon(\Sigma)$ .

Denote by  $\mathcal{M}_\varepsilon$  the set of n.f.  $\varrho_\varepsilon(f)$ ,  $f(\cdot) \in \Sigma$ , represented as

$$\varrho_\varepsilon(f) = \begin{cases} a_\varepsilon & f(\cdot) \in \Sigma_0, \\ b_\varepsilon & f(\cdot) \in \Sigma \setminus \Sigma_0. \end{cases} \quad (1.6)$$

For  $\forall \varrho_\varepsilon(\cdot) \in \mathcal{M}_\varepsilon$  introduce the risk

$$R_\varepsilon^{(a)} \left( \tilde{f}_\varepsilon, \Sigma, \varrho_\varepsilon(\cdot) \right) = \sup_{f(\cdot) \in \Sigma} \mathbf{E}_f^\varepsilon \left\{ \varrho_\varepsilon(\cdot)^{-1} \|\tilde{f}_\varepsilon(\cdot) - f(\cdot)\|_p \right\}^q.$$

Note that  $\hat{\varrho}_\varepsilon(f) \in \mathcal{M}_\varepsilon$  and  $R_\varepsilon^{(a)} \left( \tilde{f}_\varepsilon, \Sigma \right) = R_\varepsilon^{(a)} \left( \tilde{f}_\varepsilon, \Sigma, \hat{\varrho}_\varepsilon(\cdot) \right)$ . The following two statements are the simplest consequence of the fact that  $\varphi_\varepsilon(\Sigma)$  and  $\varphi_\varepsilon(\Sigma_0)$  are MRC on the sets  $\Sigma$  and  $\Sigma_0$  respectively.

For  $\forall \varrho_\varepsilon(\cdot) \in \mathcal{M}_\varepsilon$  such that  $b_\varepsilon/\varphi_\varepsilon(\Sigma) \rightarrow 0$ , as  $\varepsilon \rightarrow 0$

$$\liminf_{\varepsilon \rightarrow 0} \inf_{\tilde{f}_\varepsilon} R_\varepsilon^{(a)} \left( \tilde{f}_\varepsilon, \Sigma, \varrho_\varepsilon(\cdot) \right) = +\infty$$

For  $\forall \varrho_\varepsilon(\cdot) \in \mathcal{M}_\varepsilon$  such that  $a_\varepsilon/\varphi_\varepsilon(\Sigma_0) \rightarrow 0$ , as  $\varepsilon \rightarrow 0$

$$\liminf_{\varepsilon \rightarrow 0} \inf_{\tilde{f}_\varepsilon} R_\varepsilon^{(a)}(\tilde{f}_\varepsilon, \Sigma, \varrho_\varepsilon(\cdot)) = +\infty$$

Two last results together with (1.5) allow to make the conclusion, that n.f.  $\hat{\varrho}_\varepsilon(\cdot)$  is optimal among n.f.'s of type (1.6).

Unfortunately, even having constructed an adaptive estimator  $f_\varepsilon^{(a)}(\cdot)$ , satisfying (1.5), one can say nothing on its accuracy of estimation. It is because, the n.f.  $\hat{\varrho}_\varepsilon(f)$ , describing this accuracy, depends on an estimated function and, therefore, one is unknown. More exactly, n.f.  $\hat{\varrho}_\varepsilon(\cdot)$  depends on information whether function  $f(\cdot)$  belongs to the set  $\Sigma_0$  or not. Since such sort of information can not be obtained from noisy datas exactly, one can only state the "theoretical optimality" of the estimator  $f_\varepsilon^{(a)}(\cdot)$ , which follows from mentioned above optimality of n.f.  $\hat{\varrho}_\varepsilon(\cdot)$ . On the other hand, it seems reasonable to test the hypothesis  $H_0$  (hypothesis on belonging of an estimated function to the set  $\Sigma_0$ ) and then to use obtained results for the construction of estimators and for study their properties, in particular, the accuracy of estimation. It is evident, a result of a testing is random, one may be true, may be not, but we might expect the receiving some additional information the use of which would be natural.

These reasonings lead us to the following idea (second possibility) how to improve accuracy of estimation. This idea consists in the consideration of minimax risks with random normalizing factors of special type.

**1.2. Minimax risks with random normalizing factor.** By analogy with (1.6), for  $\forall \varepsilon \in (0, 1)$  let us consider the family of bounded, measurable w.r.t.  $X^\varepsilon$  random variables  $\varrho_\varepsilon$ , taking two values  $\{\varphi_\varepsilon(\Sigma)\}$  and  $\{a_\varepsilon\}$ , where  $0 < a_\varepsilon < \varphi_\varepsilon(\Sigma)$ . For every such  $\varrho_\varepsilon$  and for every estimator  $\tilde{f}_\varepsilon(\cdot)$  introduce the risk

$$R_\varepsilon^{(r)}(\tilde{f}_\varepsilon, \Sigma, \varrho_\varepsilon) = \sup_{f(\cdot) \in \Sigma} \mathbf{E}_f^\varepsilon \left\{ \varrho_\varepsilon^{-1} \|\tilde{f}_\varepsilon(\cdot) - f(\cdot)\|_p \right\}^q. \quad (1.7)$$

Note, that considering risks of type (1.7), we understand an improvement of accuracy of estimation as the fulfillment of the event  $\{\varrho_\varepsilon \neq \varphi_\varepsilon(\Sigma)\}$ . Since it is random event, we have to be sure, that one holds in somehow sense. Otherwise, the following artificial example

$$\varrho_\varepsilon = \begin{cases} 0 & \emptyset \\ \varphi_\varepsilon(\Sigma) & \textit{otherwise} \end{cases}$$

shows the formality of the notion "improvement". We had no such difficulties under consideration of the adaptive risk (1.4), since  $\hat{\varrho}_\varepsilon(f) = \varphi_\varepsilon(\Sigma_0) \ll \varphi_\varepsilon(\Sigma)$  for  $\forall f(\cdot) \in \Sigma_0$ . Remember, however, that theoretical possibility to improve the accuracy of estimation is connected with the acceptance of the hypothesis  $H_0$ . So, let us demand, that event  $\{\varrho_\varepsilon \neq \varphi_\varepsilon(\Sigma)\}$  holds for functions, belonging to the set  $\Sigma_0$ . In order to give these words mathematical sense, introduce the following subfamily  $\Omega_\varepsilon$  of random normalizing factors (r.n.f.). Let  $0 < \delta < 1$  be some given number and let  $\alpha_\varepsilon$ ,  $\varepsilon \in (0, 1)$  be some fixed function such that  $0 < \alpha_\varepsilon \leq 1 - \delta$  for  $\forall \varepsilon \in (0, 1)$ .

We will say that r.n.f.  $\varrho_\varepsilon \in \Omega_\varepsilon$  if

$$\limsup_{\varepsilon \rightarrow 0} \alpha_\varepsilon^{-1} \sup_{f(\cdot) \in \Sigma_0} \mathbf{P}_f^\varepsilon \{\varrho_\varepsilon = \varphi_\varepsilon(\Sigma)\} \leq 1.$$

Note, that we do not require  $\alpha_\varepsilon \rightarrow 0$  as  $\varepsilon \rightarrow 0$ , in particular one can take  $\alpha_\varepsilon = \alpha > 0$  for  $\forall \varepsilon \in (0, 1)$ . In fact  $\Omega_\varepsilon = \Omega_\varepsilon(\alpha_\varepsilon)$ , but we will omit the dependence on  $\alpha_{(\cdot)}$  in the notation, because further all statements are formulated for an arbitrary but fixed  $\alpha_{(\cdot)}$ . The sense of the condition  $\varrho_\varepsilon \in \Omega_\varepsilon$  is rather natural. One means that for  $\forall f(\cdot) \in \Sigma_0$  and for  $\forall \varrho_\varepsilon \in \Omega_\varepsilon$

$$\mathbf{P}_f^\varepsilon \{\varrho_\varepsilon \neq \varphi_\varepsilon(\Sigma)\} \geq 1 - \alpha_\varepsilon \geq 1 - \delta > 0$$

for all enough small  $\varepsilon \in (0, 1)$ . The last inequality means that the probability of the "improvement of accuracy of estimation" is positive, at least, for each function, belonging to the set  $\Sigma_0$ . On first look, it seems reasonable to choose  $\alpha_\varepsilon$ , tending to zero,  $\varepsilon \rightarrow 0$ , as fast as possible. It would guarantee

$$\mathbf{P}_f^\varepsilon \{\varrho_\varepsilon \neq \varphi_\varepsilon(\Sigma)\} \rightarrow 1, \quad \varepsilon \rightarrow 0, \quad \forall f(\cdot) \in \Sigma_0.$$

However, as we will see later on, the choice of  $\alpha_\varepsilon$  is delicate problem, and at a moment we require only  $0 < \alpha_\varepsilon < 1 - \delta$ . Note also, that the introduction of the subfamily  $\Omega_\varepsilon$  allows us to define an optimal r.n.f. and asymptotically efficient w.r.t. risk (1.7) estimator.

Let  $0 < a_\varepsilon < \varphi_\varepsilon(\Sigma)$ ,  $\varepsilon \in (0, 1)$ , be some fixed function. Denote by  $\mathcal{F}(a_{(\cdot)})$  the set of function  $b_\varepsilon$ ,  $\varepsilon \in (0, 1)$ , such that

$$\lim_{\varepsilon \rightarrow 0} \frac{b_\varepsilon}{a_\varepsilon} = 0,$$

and put

$$\Omega_\varepsilon(a_{(\cdot)}) = \left\{ \varrho_\varepsilon = (\{\varphi_\varepsilon(\Sigma)\}, \{b_\varepsilon\}) \in \Omega_\varepsilon : b_{(\cdot)} \in \mathcal{F}(a_{(\cdot)}) \right\}.$$

Thus,  $\Omega_\varepsilon(a_{(\cdot)})$  consists of the r.n.f., belonging to the family  $\Omega_\varepsilon$ , and, having second value, which is "better in order" than some given function  $a_{(\cdot)}$ .

**Definition 1.** *The r.n.f.  $\hat{\varrho}_\varepsilon = (\{\varphi_\varepsilon(\Sigma)\}, \{\varphi_\varepsilon(\alpha_{(\cdot)})\}) \in \Omega_\varepsilon$  is called optimal (asymptotically optimal) if*

1. *There exist an estimator  $f_\varepsilon^*(\cdot)$  and a constant  $M^* < \infty$ , independent on  $\varepsilon \in (0, 1)$  and  $\alpha_{(\cdot)}$ , such that*

$$\limsup_{\varepsilon \rightarrow 0} R_\varepsilon^{(r)}(f_\varepsilon^*, \Sigma, \hat{\varrho}_\varepsilon) \leq M^*. \quad (1.8)$$

2. *For  $\forall \varrho_\varepsilon \in \Omega_\varepsilon(\varphi_\varepsilon(\alpha_{(\cdot)}))$*

$$\liminf_{\varepsilon \rightarrow 0} \inf_{\tilde{f}_\varepsilon} R_\varepsilon^{(r)}(\tilde{f}_\varepsilon, \Sigma, \varrho_\varepsilon) = +\infty, \quad (1.9)$$

where  $\inf$  is taken over all possible estimators.

**Definition 2.** *Let  $\hat{\varrho}_\varepsilon$  be an optimal r.n.f. Then an estimator  $f_\varepsilon^*(\cdot)$ , satisfying (1.8), is called  $\alpha_{(\cdot)}$ -adaptive.*

**Remark 1.** *As it follows from (1.9), the function  $\varphi_\varepsilon(\alpha_{(\cdot)})$ , being the second value of optimal r.n.f.  $\hat{\varrho}_\varepsilon$  can not be improved in order. The first value  $\varphi_\varepsilon(\Sigma)$  can not be improved in order as well, because one is MRC. Both these facts together with (1.8) explain why  $\hat{\varrho}_\varepsilon$  is called optimal.*

**Remark 2.** *By definition  $\hat{\varrho}_\varepsilon \leq \varphi_\varepsilon(\Sigma)$  for  $\forall \varepsilon \in (0, 1)$  and for any  $\alpha_{(\cdot)}$ . Therefore, from (1.8), any  $\alpha_{(\cdot)}$ -adaptive estimator is a.e. estimator on the set  $\Sigma$  w.r.t. the risk (1.1). It means, that considering risks of type (1.7), we, in fact, do not leave the frameworks of the standard minimax approach.*

**Remark 3.** *We know only that by definition  $\varphi_\varepsilon(\alpha_{(\cdot)}) < \varphi_\varepsilon(\Sigma)$ . So, if  $\varphi_\varepsilon(\alpha_{(\cdot)}) \asymp \varphi_\varepsilon(\Sigma)$ , as  $\varepsilon \rightarrow 0$ , then from (1.9) any improvement of the accuracy of estimation is impossible in this case. Such statistical setups exist, f.e., it is typical for cases, when minimax risk is described by uniform norm ( $p = \infty$ ). However, the study of such kind of problems lies beyond the scope of the paper. We refer to the recent paper of Low (1996), where similar results for the case  $p = \infty$  were obtained. Certainly, the case  $\varphi_\varepsilon(\alpha_{(\cdot)}) = o(\varphi_\varepsilon(\Sigma))$ , as  $\varepsilon \rightarrow 0$ , is much more interesting for our purposes, and only such setups are studied later on.*

Another interesting question is: what is a connection between  $\varphi_\varepsilon(\alpha_{(\cdot)})$  and  $\varphi_\varepsilon(\Sigma_0)$ ? It is intuitively clear, that  $\varphi_\varepsilon(\alpha_{(\cdot)})$  can not be better in order than  $\varphi_\varepsilon(\Sigma_0)$ . The exact statement is given by the following Proposition 1.

**Proposition 1.** *Let  $\hat{\varrho}_\varepsilon = (\{\varphi_\varepsilon(\Sigma)\}, \{\varphi_\varepsilon(\alpha_{(\cdot)})\})$  be some optimal r.n.f. Suppose, there exist an estimator  $\bar{f}_\varepsilon(\cdot)$  such that for some  $q_1 > q$*

$$\limsup_{\varepsilon \rightarrow 0} \sup_{f(\cdot) \in \Sigma_0} \mathbf{E}_f^\varepsilon \left\{ \varphi_\varepsilon^{-1}(\Sigma_0) \|\bar{f}_\varepsilon(\cdot) - f(\cdot)\|_p \right\}^{q_1} \leq R < \infty.$$

*Then for  $\forall \alpha_{(\cdot)}$  such that  $0 < \alpha_\varepsilon \leq \alpha_0$ ,  $\varepsilon \in (0, 1)$ ,*

$$\liminf_{\varepsilon \rightarrow 0} \left( \frac{\varphi_\varepsilon(\alpha_{(\cdot)})}{\varphi_\varepsilon(\Sigma_0)} \right) \geq \left( \frac{l_0}{2M^*} \right)^{\frac{1}{q}},$$

*where*

$$0 < l_0 = \liminf_{\varepsilon \rightarrow 0} \inf_{\bar{f}_\varepsilon} R_\varepsilon(\bar{f}_\varepsilon, \Sigma_0, \varphi_\varepsilon(\Sigma_0)); \quad \alpha_0 = (2^{-1}l_0)^{\frac{q_1}{q_1-q}} R^{-\frac{q}{q_1-q}}.$$

**Remark 4.** *The assumption of the proposition means that  $\varphi_\varepsilon(\Sigma_0)$  is MRC on the set  $\Sigma_0$  not only for the loss function  $|\cdot|^q$ , but for the loss function  $|\cdot|^{q_1}$  as well. It is typical for the asymptotical statistics, that one and the same function is MRC for wide class of loss functions, see, f.e., Ibragimov, Khasminskii (1981). Note also that  $l_0 > 0$  by definition of MRC.*

## 2. Application to the construction of confidence sets

In this section we show how to apply the notions of optimal r.n.f. and  $\alpha_{(\cdot)}$ -adaptive estimator to the construction of confidence sets. Fix some  $0 < \gamma < 1$  and let  $\hat{\varrho}_\varepsilon$  and  $f_\varepsilon^*(\cdot)$  be the optimal r.n.f. and  $\alpha_{(\cdot)}$ -adaptive estimator respectively. The

function  $\alpha_{(\cdot)}$  is supposed to be chosen on an arbitrary way. Then, from (1.8) and Markov inequality, one obtains for all small enough  $\varepsilon > 0$

$$\sup_{f(\cdot) \in \Sigma} \mathbf{P}_f^\varepsilon \left\{ \hat{\varrho}_\varepsilon^{-1} \|f_\varepsilon^*(\cdot) - f(\cdot)\|_p > \frac{M^*}{\gamma} \right\} \leq \gamma.$$

This is equivalent

$$\mathbf{P}_f^\varepsilon \left\{ \|f_\varepsilon^*(\cdot) - f(\cdot)\|_p \leq \frac{M^*}{\gamma} \hat{\varrho}_\varepsilon \right\} \geq 1 - \gamma.$$

uniformly w.r.t.  $f(\cdot) \in \Sigma$ . It means, that with given probability  $1 - \gamma$  an estimated function lies inside the  $L_p$ -ball, with center in the "point"  $f_\varepsilon^*(\cdot)$  and of the "radius"  $\frac{M^*}{\gamma} \hat{\varrho}_\varepsilon$ . By definition, pair  $(f_\varepsilon^*(\cdot), \hat{\varrho}_\varepsilon)$  is computable by observation  $X^\varepsilon$  (measurable w.r.t.  $X^\varepsilon$ ), and, therefore, if event  $\{\hat{\varrho}_\varepsilon = \varphi_\varepsilon(\alpha_{(\cdot)})\}$  holds, we guarantee essentially more precise coverage of an estimated function.

### 3. Relations to the adaptive estimation and to the hypothesis testing

In this section we briefly discuss the relations between the problem of finding of an optimal r.n.f. and an  $\alpha_{(\cdot)}$ -adaptive estimator and problems, arising in the adaptive estimation and in the hypothesis testing.

Let  $\hat{f}_\varepsilon(\cdot)$  and  $\hat{f}_\varepsilon^{(0)}(\cdot)$  be a.e. estimators on the sets  $\Sigma$  and  $\Sigma_0$  respectively. Further we will see that  $\alpha_{(\cdot)}$ -adaptive estimator  $f_\varepsilon^*(\cdot)$  can be often represented as follows.

$$f_\varepsilon^*(\cdot) = \begin{cases} \hat{f}_\varepsilon^{(0)}(\cdot) & \hat{\varrho}_\varepsilon = \varphi_\varepsilon(\alpha_{(\cdot)}) \\ \hat{f}_\varepsilon(\cdot) & \hat{\varrho}_\varepsilon = \varphi_\varepsilon(\Sigma) \end{cases} \quad (3.1)$$

**Proposition 2.** *Let  $\alpha_\varepsilon = \mathcal{O}(\{\varphi_\varepsilon(\Sigma_0)\}^q)$ , as  $\varepsilon \rightarrow 0$  and let  $f_\varepsilon^*(\cdot)$  be  $\alpha_\varepsilon$ -adaptive estimator, represented by (3.1).*

*Then  $f_\varepsilon^*(\cdot)$  is adaptive estimator, i.e risk (1.5) (or (1.2)) of this estimator is finite.*

As it follows from proposition 2, if  $\alpha_\varepsilon$  tends to zero rather quickly, then  $\alpha_{(\cdot)}$ -adaptive estimator is, at the same time, adaptive estimator.

Let us now consider the relations to the hypotheses testing problems. The fulfillment of the event  $\{\hat{\varrho}_\varepsilon = \varphi_\varepsilon(\alpha_{(\cdot)})\}$  can be treated as the acceptance of the hypothesis  $\mathbf{H}_0 : f(\cdot) \in \Sigma_0$ . Then, the assumption  $\hat{\varrho}_\varepsilon \in \Omega_\varepsilon$  means that first type error probability is bounded by  $\alpha_{(\cdot)}$ .

**Proposition 3.** *Let  $\hat{\varrho}_\varepsilon$  be optimal r.n.f. and let  $\alpha_{(\cdot)}$ -adaptive estimator  $f_\varepsilon^*(\cdot)$  be represented by (3.1). Suppose, also, that  $\hat{f}_\varepsilon^{(0)}(\cdot) \in \Sigma_0$  for  $\forall \varepsilon \in (0, 1)$ .*

*Then for  $\forall 0 < \gamma < 1 \exists H_\gamma > 0$  such that*

$$\limsup_{\varepsilon \rightarrow 0} \sup_{f(\cdot) \in \Sigma_\varepsilon(H_\gamma)} \mathbf{P}_f^\varepsilon \left\{ \hat{\varrho}_\varepsilon = \varphi_\varepsilon(\alpha_{(\cdot)}) \right\} \leq \gamma,$$

where for  $\forall H > 0$

$$\Sigma_\varepsilon(H) = \left\{ f(\cdot) \in \Sigma : \inf_{f_0 \in \Sigma_0} \|f(\cdot) - f_0(\cdot)\|_p \geq H \varphi_\varepsilon(\alpha_{(\cdot)}) \right\}.$$



The statement of Proposition 3 means, that hypothesis  $H_0$  can be tested versus the family of local alternative  $\mathbf{H}_\varepsilon : f(\cdot) \in \Sigma_\varepsilon(H_\gamma)$  with prescribed 1-st and 2-d type errors probabilities.

#### 4. Examples of the statistical models and of the hypotheses, corresponding to them

In this section we consider 3 particular models: white Gaussian noise model, multivariate regression model and probability density model. For each model we discuss several hypotheses  $H_0$ , which is seemed naturally to investigate in context of the problems, presented in the paper. All these hypotheses possess the following property: more precise estimation procedures are available under them. We also want to mention that all mathematical results, presented in the examples, are valid under some additional assumptions. We will not describe and discuss them, because it is not required for further consideration, and we give only references on the papers, where exact results can be found.

**4.1. White Gaussian Noise Model.** Let statistical experiment be generated by the observation  $X^\varepsilon$ , which is the sample path of the stochastic process  $X_\varepsilon(\cdot)$ , satisfying on the interval  $[0, 1]$  the stochastic differential equation

$$dX_\varepsilon(t) = f(t)dt + \varepsilon dw(t),$$

where  $w(t)$  is standard Wiener process. Thus,  $X^\varepsilon = (X_\varepsilon(t), 0 \leq t \leq 1)$ . Remind, that  $\varepsilon$  is the small parameter, and the case  $\varepsilon \rightarrow 0$  is of our interest. Let  $\beta > 0$ ,  $Q > 0$  be some given constants, and let  $\beta = m + \tau$ , where  $m \geq 0$  is integer and  $0 < \tau \leq 1$ . Let  $H(\beta, Q)$  be Holder space, i.e. the set of  $m$ -times continuously differentiable functions  $f(\cdot)$  whose  $m$ -th derivative, satisfies on  $[0, 1]$  Holder condition with exponent  $\tau$  and constant  $Q$ , i.e.

$$|f^{(m)}(t_1) - f^{(m)}(t_2)| \leq Q|t_1 - t_2|^\tau, \quad \forall t_1, t_2 \in [0, 1].$$

Here  $f^{(m)}(\cdot)$  denotes  $m$ -th derivative of  $f(\cdot)$ . In this model we consider univariate case, hence  $s = 1$ ,  $\|\cdot\|_p$ ,  $1 \leq p < \infty$  and  $\|\cdot\|_\infty$  are usual  $L_p$ -norm or C-norm respectively, which are determined on the  $[0, 1]$ .

Let  $\Sigma = H(\beta, Q) \cap \Sigma_1(L)$  for some given  $\beta, Q$  and  $L$ . It is well-known (Ibragimov and Khasminskii (1980)) that

$$\varphi_\varepsilon(\Sigma) = \begin{cases} \varepsilon^{\frac{2\beta}{2\beta+1}} & 1 \leq p < \infty, \\ \left(\varepsilon \sqrt{\ln \frac{1}{\varepsilon}}\right)^{\frac{2\beta}{2\beta+1}} & p = \infty. \end{cases}$$

This rate is attained by linear estimators, f.e., by kernel one with properly chosen kernel and bandwidth. Other classes of smooth functions, such as Sobolev and Besov ones, can be used for the description of the set  $\Sigma$  as well. Now let us consider some possible hypotheses.

4.1.1. *Hypothesis on parametric subfamily.* Consider the following hypothesis

$$\mathbf{H}_0 : f(\cdot) \in \Sigma_0 = \{f(t) = f_0(t, \vartheta), \forall t \in [0, 1], \vartheta \in \Theta \subset \mathbf{R}^s\},$$

where function  $f_0(\cdot, \cdot)$ , set  $\Theta$  and integer  $s \geq 1$  are given. Under some regularity assumptions, Ibragimov and Khasminskii (1981, ch.2) the MRC on the set  $\Sigma_0$  is  $\varphi_\varepsilon(\Sigma_0) = \varepsilon$  and a.e. estimator can be constructed as  $\hat{f}_\varepsilon^{(0)}(\cdot) = f_0(\cdot, \hat{\vartheta}_\varepsilon)$ , where  $\hat{\vartheta}_\varepsilon$  is the maximum likelihood estimator. Note that  $\varepsilon = \varphi_\varepsilon(\Sigma_0) \ll \varphi_\varepsilon(\Sigma) = \varepsilon^{\frac{2\beta}{2\beta+1}}$  ( $p < \infty$ ) or  $(\varepsilon \sqrt{\ln \frac{1}{\varepsilon}})^{\frac{2\beta}{2\beta+1}}$  ( $p = \infty$ ) for all  $\beta \in (0, \infty)$ . One of the classical examples of parametric subfamily is yielded polynomial regression, i.e.

$$f_0(t, \vartheta) = \sum_{i=1}^s \vartheta_i t^{i-1}, \quad (\vartheta_1, \dots, \vartheta_s) \in \Theta \subset \mathbf{R}^s.$$

4.1.2. *Hypothesis on smoothness.* Let

$$\mathbf{H}_0 : H(\gamma, P) \cap \Sigma,$$

where  $\gamma > \beta$  and  $H(\gamma, P), P > 0$ , is another Holder space. In this case the set  $\Sigma_0$  consists of the functions, which are smoother than functions, belonging to the set  $\Sigma$ . The MRC  $\varphi_\varepsilon(\Sigma_0)$  is given by formulae

$$\varphi_\varepsilon(\Sigma_0) = \begin{cases} \varepsilon^{\frac{2\gamma}{2\gamma+1}} & 1 \leq p < \infty, \\ \left(\varepsilon \sqrt{\ln \frac{1}{\varepsilon}}\right)^{\frac{2\gamma}{2\gamma+1}} & p = \infty, . \end{cases}$$

and we see that again  $\varphi_\varepsilon(\Sigma_0) \ll \varphi_\varepsilon(\Sigma)$ .

For the sets  $\Sigma_0$ , described in examples 4.1.1 and 4.1.2, the method of the construction of adaptive estimators is the simple application of the results, obtained in Lepskii (1991), for all  $1 \leq p \leq \infty$ ,  $0 < \beta < \gamma < \infty$  and  $q > 0$ .

In the section 5, for hypotheses, presented in examples 4.1.1 and 4.1.2, we find optimal r.n.f. and construct  $\alpha_{(\cdot)}$ -adaptive estimator.

4.2. **Multivariate regression.** Let statistical experiment be generated by the observation  $X^n$  (here and in the next example  $\frac{1}{n}$  plays role  $\varepsilon$ ), obtained in the multivariate regression model, i.e.  $X^n = \{(y_1, Z_1), \dots, (y_n, Z_n)\}$ , where

$$y_i = f(Z_i) + \varepsilon_i, \quad i = 1, \dots, n.$$

Here  $Z_i = (z_i^{(1)}, \dots, z_i^{(s)})$ ,  $i = 1, \dots, n$ , are i.i.d random vectors with common probability density  $p(\cdot)$ , defined on the unit cube  $[0, 1]^s$ ;  $\varepsilon_i$ ,  $i = 1, \dots, n$ , are i.i.d. random variables,  $\mathbf{E}\varepsilon_1 = 0$ ,  $\mathbf{E}\varepsilon_1^2 = \sigma^2 < \infty$ .

Let  $H_s(\beta, Q)$ ,  $\beta = m + \tau$ ,  $Q > 0$ , be isotropic Holder space on the unit cube  $[0, 1]^s$ . There are several equivalent definitions of isotropic (anisotropic) Holder (or Sobolev and Besov) spaces, see, f.e., Nikolskii (1975). We will use the following one. Fix some  $i \in \{1, 2, \dots, s\}$  and denote  $f_i^{(m)}(Z) = \frac{\partial^m f(Z)}{\partial z_i^m}$ . Put  $Z^{(l)} = (z_1^{(l)}, \dots, z_i^{(l)}, \dots, z_s^{(l)})$ ,  $l = 1, 2$ . We say that function  $f(\cdot)$  belongs to the Holder space  $H_s(\beta, Q)$ , if

$$\sup_{(z_1, \dots, z_{i-1}, z_{i+1}, \dots, z_s) \in [0, 1]^{s-1}} \left| f_i^{(m)}(Z^{(1)}) - f_i^{(m)}(Z^{(2)}) \right| \leq Q \left| z_i^{(1)} - z_i^{(2)} \right|^\tau,$$

for  $\forall i = 1, \dots, s$  and  $\forall z_i^{(1)}, z_i^{(2)} \in [0, 1]$ .

Roughly speaking, for each fixed  $(z_1, \dots, z_{i-1}, z_{i+1}, \dots, z_s) \in [0, 1]^{s-1}$  functions  $g_f^i(z) = f(z_1, \dots, z_{i-1}, z, z_{i+1}, \dots, z_s)$  belong, as function of  $z$ , to the Holder space  $H(\beta, Q)$  on the interval  $[0, 1]$  for  $\forall i = 1, \dots, s$ . In other words, any function  $f(\cdot) \in H_s(\beta, Q)$  has one and the same smoothness  $\beta = m + \tau$  in each direction.

Let  $\Sigma = H_s(\beta, Q) \cap \Sigma_s(L)$ , for some given  $\beta, Q, L$  and  $s$ . Put also  $p = 2$ , i.e. we will consider the losses, being  $L_2$ -norm on the unit cube  $[0, 1]^s$ . The MRC is given by formulae  $\varphi_n(\Sigma) = n^{-\frac{\beta}{2\beta+s}}$ , which is obtained by Nussbaum (1986).

Now let us describe some possible hypotheses. Certainly, the hypothesis on parametric subfamily and the hypothesis on smoothness are of interest in this case as well. However, as it seems to us, hypotheses, which could be called "hypotheses on structure", are more important under consideration of multidimensional statistical models.

**4.2.1. Dimensionality hypothesis.** This hypothesis consists in the assumption that a regression function actually depends on  $t < s$  significant variables. Thus, formally,

$$\mathbf{H}_0 : \exists 1 \leq t < s, i_1, \dots, i_t, F : \mathbf{R}^t \rightarrow \mathbf{R}^1 : f(z_1, \dots, z_s) = F(z_{i_1}, \dots, z_{i_t}).$$

Evidently, that the implication  $f(\cdot) \in H_s(\beta, Q) \implies F(\cdot) \in H_t(\beta, Q)$  takes place, and, hence,  $\varphi_n(\Sigma_o) = n^{-\frac{\beta}{2\beta+t}}$ , and we see that  $\varphi_n(\Sigma_o) \ll \varphi_n(\Sigma)$ , for  $\forall 1 \leq t < s$ .

**4.2.2. Hypothesis on additive structure.** This hypothesis consists in the assumption, that multivariate regression function can be represented as the sum of univariate functions.

$$\mathbf{H}_0 : \exists f_k : \mathbf{R}^1 \rightarrow \mathbf{R}^1, k = 1, \dots, s : f(z_1, \dots, z_s) = \sum_{k=1}^s f_k(z_k).$$

A lot of papers are devoted to the estimation problem under additivity hypothesis. It is because, the MRC under this hypothesis coincides with MRC for univariate case, i.e.

$$\varphi_n(\Sigma_0) = n^{-\frac{\beta}{2\beta+1}} \ll \varphi(\Sigma).$$

Apparently, the first paper, where this result has been obtained, is Stone (1985). We also mention the paper of Linton and Nielsen (1995), where the same result has been proved under rather mild assumptions.

It is also reasonable to combine the hypothesis on additive structure with the hypothesis on smoothness.

$$\tilde{\mathbf{H}}_0 : \exists f_k(\cdot) \in H(\gamma, P), k = 1, \dots, s, \gamma > \beta : f(z_1, \dots, z_s) = \sum_{k=1}^s f_k(z_k).$$

By the same reasons, under this hypothesis

$$\varphi_n(\Sigma_0) = n^{-\frac{\gamma}{2\gamma+1}}.$$

4.2.3. *Hypothesis on single index structure.* This hypothesis consists in the assumption that there exists some direction, where a multivariate function is an univariate one.

$$\mathbf{H}_0 : \exists \vartheta \in \mathbf{R}^s, \|\vartheta\| = 1, \text{ and } F : \mathbf{R}^1 \rightarrow \mathbf{R}^1 : f(z_1, \dots, z_s) = F\left(\sum_{k=1}^s \vartheta_k z_k\right).$$

As it follows, f.e., from Speckman (1988), the MRC under this hypothesis is the same as for the univariate case,

$$\varphi_n(\Sigma_0) = n^{-\frac{\beta}{2\beta+1}},$$

and we see, that  $\varphi_n(\Sigma_0) \ll \varphi_n(\Sigma)$ .

It is also possible to combine the hypothesis on single index structure with the hypothesis on parametric subfamily. To do this, it is enough to suppose that function  $F(\cdot)$  is known. For example, putting  $F(x) = x, \forall x \in \mathbf{R}^1$ , we arrive to the "linearity" hypothesis

$$\tilde{\mathbf{H}}_0 : f(\cdot) \in \Sigma_0 = \{f(\cdot) : f(z) = \vartheta_1 z_1 + \dots + \vartheta_s z_s\}.$$

It is clear, that under this hypothesis

$$\varphi_n(\Sigma_0) = n^{-\frac{1}{2}}.$$

4.3. **Probability density estimation.** Let statistical experiment be generated by the observation  $X^n = (X_1, \dots, X_n)$ , where  $X_i = (X_i^{(1)}, \dots, X_i^{(s)})$ ,  $i = 1, \dots, n$ , are i.i.d. random vectors with common probability density  $f(x_1, \dots, x_s)$ . Let  $f(\cdot) \in \Sigma$ , where, as in example 4.2.,  $\Sigma = H_s(\beta, Q) \cap \Sigma_s(L)$  for some fixed  $\beta, Q, L$  and  $s$ . Let again  $p = 2$ . Then, Nussbaum (1986), the MRC on the set  $\Sigma$  is

$$\varphi_n(\Sigma) = n^{-\frac{\beta}{2\beta+2s}}.$$

4.3.1. *Hypothesis of independence.* This hypothesis is classical in the theory of hypothesis testing. One consists in the assumption that for  $\forall i = \overline{1, n}$  the components of the vector  $X_i = (X_i^{(1)}, \dots, X_i^{(s)})$  are independent random variables.

$$\mathbf{H}_0 : \exists f_k(\cdot) : \mathbf{R}^1 \rightarrow \mathbf{R}^1, \int_{-\infty}^{\infty} f_k(x) dx = 1, k = \overline{1, s} : f(x_1, \dots, x_s) = \prod_{k=1}^s f_k(x_k).$$

Under this hypothesis, for  $\forall k = \overline{1, s}$  each univariate density  $f_k(\cdot)$  can be estimated separately, using only the corresponding observations  $X_1^{(k)}, X_2^{(k)}, \dots, X_n^{(k)}$ . Let  $\hat{f}_k(\cdot) = \hat{f}_k(\cdot; X_1^{(k)}, X_2^{(k)}, \dots, X_n^{(k)})$  be, for example, Nadaraya-Watson estimator, providing univariate MRC -  $n^{-\frac{\beta}{2\beta+1}}$ . Under hypothesis  $H_0$   $\hat{f}_k(\cdot)$  are i.i.d. random variables, and, therefore, the estimator  $\hat{f}^{(0)}(\cdot) = \prod_{k=1}^s \hat{f}_k(\cdot)$  provides univariate MRC of estimation for  $f(x_1, \dots, x_s) = \prod_{k=1}^s f_k(x_k)$ . Thus,

$$\varphi_n(\Sigma_0) = n^{-\frac{\beta}{2\beta+1}}.$$

As regards to the problem of adaptive estimation for multivariate statistical models, there is much less known in comparison with univariate ones. In the case of isotropic spaces, apparently the general theory, developed in Lepski (1991, 1992a),

could be applied. In the case of anisotropic spaces we know only recent papers of Neumann(1995) and Birge and Massart (1995), concerning with adaptive estimation over scales of anisotropic Besov spaces and of anisotropic Holder spaces respectively.

The consideration of minimax risks with r.n.f. for multidimensional models is the subject of a series of separated papers. Here we would like only to present one conjecture in this direction.

**Conjecture 1.** *Let  $\Sigma = H_s(\beta, Q) \cap \Sigma_s(L)$  and let we believe the hypothesis on additive structure or the hypothesis on single index structure. Then the function  $\varphi_\varepsilon(\alpha_{(\cdot)})$ , being the second value of optimal r.n.f., and which is understood as the "improvement of accuracy of estimation", is given by the following formulae.*

$$\varphi_n(\alpha_n) = \left( n \sqrt{\ln \frac{1}{\alpha_n}} \right)^{-\frac{2\beta}{4\beta+s}}.$$

As we see, if our conjecture is true, then the "improvement" always exists, and for the dimensionality  $s = 2$  this "improvement" differs from MRC under hypothesis only by factor  $\left( \ln \frac{1}{\alpha_n} \right)^{\frac{\beta}{4\beta+2}}$ . In particular, if  $\alpha_n \equiv \alpha$ , then the function  $\varphi_n(\alpha_{(\cdot)})$  coincide with MRC on hypothesis set, and, due to Proposition 1, this is the best possible improvement of the accuracy of estimation.

## 5. White Gaussian Noise Model

Let we observe the sample path of the stochastic process  $X_\varepsilon(\cdot)$ , satisfying on the interval  $[0, 1]$  the stochastic differential equation

$$dX_\varepsilon(t) = f(t)dt + \varepsilon dw(t), \quad (5.1)$$

where  $w(\cdot)$  is standard Wiener process,  $\varepsilon \rightarrow 0$  is the small parameter.

Let  $\beta > 0$ ,  $Q > 0$  be some given constants, and let  $\beta = m + \tau$ , where  $m \geq 0$  is integer and  $0 < \tau \leq 1$ . Here and later we suppose that function  $f(\cdot)$ , generated the equation (5.1), belongs to the space  $\Sigma = H(\beta, Q) \cap \Sigma_1(L)$ .

**5.1. Hypothesis on parametric subfamily.** Let us suppose that we believe the following hypothesis

$$\mathbf{H}_0 : f(\cdot) \in \Sigma_0 = \{f(\cdot) \in \Sigma : f(\cdot) = f_0(t, \vartheta), \forall t \in [0, 1], \vartheta \in \Theta \subset \mathbf{R}^s\},$$

where function  $f_0(\cdot, \cdot)$ , set  $\Theta$  and  $s \geq 1$  are given.

For each integer  $l \geq 1$  and for each vector  $z \in \mathbf{R}^l$  denote  $\|z\| = \left( \sum_{i=1}^l z_i^2 \right)^{\frac{1}{2}}$  and we will omit the dependence of  $l$  in the notation of  $\|\cdot\|$ . We suppose that the following assumptions fulfill.

**A1.** The set  $\Theta$  is bounded, closed subset in  $\mathbf{R}^s$ ,  $s \geq 1$ .

**A2.** There exist  $\nu > 0$  and  $L_\nu > 0$  such that

$$\sup_{t \in [0, 1]} |f(t, \vartheta_1) - f(t, \vartheta_2)| \leq L_\nu \|\vartheta_1 - \vartheta_2\|^\nu, \text{ for } \forall \vartheta_1, \vartheta_2 \in \Theta.$$

**A3.**  $f_0(\cdot, \vartheta) \in H(\beta, Q) \cap \Sigma_1(L)$  for  $\forall \vartheta \in \Theta$ .

**A4.**  $\exists \vartheta_0 \in \Theta$  and  $\exists Q_1 < Q$  such that  $f_0(\cdot, \vartheta_0) \in H(\beta, Q_1)$ .

Put  $\varphi_\varepsilon(\beta) = \varepsilon^{\frac{2\beta}{2\beta+1}}$  and let  $\|f(\cdot)\|_2 = \left(\int_0^1 f^2(t) \mathbf{d}t\right)^{\frac{1}{2}}$  be  $L_2$ -norm on the interval  $[0, 1]$ . Remind that  $\varphi_\varepsilon(\beta)$  is MRC on the set  $H(\beta, Q)$ , in particular, when minimax risk is described by  $L_2$ -losses.

Now let us introduce the minimax risk with r.n.f., corresponding to the estimation problem on the space  $\Sigma$  and to the hypothesis  $H_0$ .

Fix some  $0 < \delta < 1$  and let  $0 < \alpha_\varepsilon < 1 - \delta$ ,  $\varepsilon \in [0, 1]$ , be some given function. Denote by  $\Omega_\varepsilon$  the family of measurable w.r.t.  $(X_\varepsilon(t), 0 \leq t \leq 1)$ , random variables, taking two values  $\varphi_\varepsilon(\beta)$  and  $a_\varepsilon$ , where  $0 < a_\varepsilon < \varphi_\varepsilon(\beta)$ , and satisfying the following inequality

$$\limsup_{\varepsilon \rightarrow 0} \alpha_\varepsilon^{-1} \sup_{\vartheta \in \Theta} \mathbf{P}_\vartheta \{\varrho_\varepsilon = \varphi_\varepsilon(\beta)\} \leq 1. \quad (5.2)$$

Here  $\mathbf{P}_\vartheta$  denotes the measure, generated by the process  $X_\varepsilon(t), 0 \leq t \leq 1$ , when function  $f(\cdot)$  in (5.1) belongs to the set  $\Sigma_0$ , i.e.  $f(\cdot) = f_0(\cdot, \vartheta)$  for some  $\vartheta \in \Theta$ . For  $\forall \varrho_\varepsilon \in \Omega_\varepsilon$  and for every an arbitrary estimator  $\tilde{f}_\varepsilon(\cdot)$  consider the risk

$$R_\varepsilon^{(r)}(\tilde{f}_\varepsilon, \Sigma, \varrho_\varepsilon) = \sup_{f(\cdot) \in \Sigma} \mathbf{E}_f \left( \varrho_\varepsilon^{-1} \|\tilde{f}_\varepsilon(\cdot) - f(\cdot)\|_2 \right)^q, \quad (5.3)$$

where  $q > 0$  is some fixed constant.

Now let us construct the r.n.f.  $\hat{\varrho}_\varepsilon \in \Omega_\varepsilon$ , being optimal, in accordance with the definition 1, and  $\alpha_{(\cdot)}$ -adaptive estimator  $f_\varepsilon^*(\cdot)$ .

Let  $\Lambda = \Lambda_m = \{\lambda\}_{i,j=\overline{1,m+1}}$  be the matrix with the elements  $\lambda_{i,j} = (i+j-1)^{-1}$ . Put for an arbitrary integer  $N$  and  $k = \overline{1, N}$ ,  $t_k = t_k(N) = k/N$ ;  $\Delta_k = [t_{k-1}, t_k)$ . Also, for  $\forall k = \overline{1, N}$  introduce the vectors  $(\mathbf{x}^{(k)})' = (x_1^{(k)}, \dots, x_{m+1}^{(k)})$  and  $(\mathbf{y}^{(k)})' = (y_1^{(k)}, \dots, y_{m+1}^{(k)})$  as

$$\begin{aligned} x_j^{(k)} &= N^j \int_{\Delta_k} (t - t_k)^{j-1} \mathbf{d}X_\varepsilon(t), \quad j = \overline{1, m+1}; \\ \mathbf{y}^{(k)} &= \Lambda^{-\frac{1}{2}} \mathbf{x}^{(k)}. \end{aligned}$$

Here sign " / " means transposition. The vectors  $\mathbf{y}^{(k)}$ ,  $k = \overline{1, N}$ , are well-defined, since matrix  $\Lambda$  is strictly positively defined for  $\forall m \geq 1$ , . For  $\forall k = \overline{1, N}$  and  $\forall t \in \Delta_k$  put  $(\mathbf{d}^{(k)}(t))' = (1, N(t - t_{k-1}), \dots, (N(t - t_{k-1}))^m)$ . Denote for  $\forall t \in [0, 1]$

$$f_N(t) = \sum_{k=1}^N (\mathbf{x}^{(k)})' \Lambda^{-1} \mathbf{d}^{(k)}(t) I\{t \in \Delta_k\} \quad (5.4)$$

or in equivalent form

$$f_N(t) = \sum_{k=1}^N (\mathbf{y}^{(k)})' \Lambda^{-\frac{1}{2}} \mathbf{d}^{(k)}(t) I\{t \in \Delta_k\}. \quad (5.5)$$

Put

$$N_\varepsilon(\beta, Q) = \left[ \left( \frac{2\beta Q}{m! \sqrt{m+1}} \right)^{\frac{2}{2\beta+1}} \varepsilon^{-\frac{2}{2\beta+1}} \right];$$

$$N_\varepsilon = N_\varepsilon(\alpha_\varepsilon) = \left[ Z_0^{-4} \left( \varepsilon^4 \ln \frac{1}{\alpha_\varepsilon} \right)^{-\frac{1}{4\beta+1}} \right],$$

where constant  $Z_0 = Z_0(\beta, m, Q)$  will be presented below. Define for  $\forall t \in [0, 1]$

$$\hat{f}_\varepsilon(t, \beta, Q) = f_{N_\varepsilon(\beta, Q)}(t); \quad \hat{f}_\varepsilon(t) = f_{N_\varepsilon}(t).$$

As it will be shown later,  $\hat{f}_\varepsilon(\cdot, \beta, Q)$  is the estimator of a function  $f(\cdot) \in H(\beta, Q)$ , constructed by method of piecewise polynomial approximation, and, providing the MRC  $\varphi_\varepsilon(\beta)$ .

For  $\forall \vartheta \in \Theta$  and for  $\forall k = \overline{1, N_\varepsilon}$  introduce the vectors  $(\mathbf{x}^{(k)}(\vartheta))' = (x_1^{(k)}(\vartheta), \dots, x_{m+1}^{(k)}(\vartheta))$  and  $(\mathbf{y}^{(k)}(\vartheta))' = (y_1^{(k)}(\vartheta), \dots, y_{m+1}^{(k)}(\vartheta))$  as follows.

$$x_j(\vartheta) = N^j \int_{\Delta_k} (t - t_{k-1})^{j-1} f_0(t, \vartheta) \mathbf{d}t, \quad j = \overline{1, m+1};$$

$$\mathbf{y}^{(k)}(\vartheta) = \Lambda^{-\frac{1}{2}} \mathbf{x}^{(k)}(\vartheta).$$

Put, also, for  $\forall \vartheta \in \Theta$  and for  $\forall t \in [0, 1]$

$$\bar{f}_\varepsilon(t, \vartheta) = \sum_{k=1}^{N_\varepsilon} (\mathbf{y}^{(k)}(\vartheta))' \Lambda^{-\frac{1}{2}} \mathbf{d}^{(k)}(t) I\{t \in \Delta_k\}.$$

Denote for  $\forall \vartheta \in \Theta$

$$d_\varepsilon(\vartheta) = \|\hat{f}_\varepsilon(\cdot) - \bar{f}_\varepsilon(\cdot, \vartheta)\|_2. \quad (5.6)$$

Simplest direct calculations show that

$$d_\varepsilon(\vartheta) = \left[ \frac{1}{N_\varepsilon} \sum_{k=1}^{N_\varepsilon} \|\mathbf{y}^{(k)} - \mathbf{y}^{(k)}(\vartheta)\|^2 \right]^{\frac{1}{2}}. \quad (5.7)$$

As it follow from the assumption **A2**, there exists independent on  $\varepsilon$  constant  $C = C(\nu, L_\nu, m)$  such that

$$|d_\varepsilon(\vartheta_1) - d_\varepsilon(\vartheta_2)| \leq C \|\vartheta_1 - \vartheta_2\|^\nu, \quad \forall \vartheta_1, \vartheta_2 \in \Theta. \quad (5.8)$$

From here and assumption **A1** there exist, see, f.e. Jennrich (1969),  $\hat{\vartheta}_\varepsilon \in \Theta$ , being measurable w.r.t.  $X_\varepsilon(t), 0 \leq t \leq 1$ , and such that

$$\hat{\vartheta}_\varepsilon = \operatorname{arginf}_{\vartheta \in \Theta} d_\varepsilon(\vartheta). \quad (5.9)$$

Denote for  $\forall \vartheta \in \Theta$

$$\hat{d}_\varepsilon(\vartheta) = d_\varepsilon^2(\vartheta) - \varepsilon^2 N_\varepsilon(m+1),$$

and let

$$\varphi_\varepsilon(\alpha_\varepsilon) = \left( \varepsilon^4 \ln \frac{1}{\alpha_\varepsilon} \right)^{\frac{\beta}{4\beta+1}},$$

$$\lambda = Z_0^{-1} (4(m+1))^{\frac{1}{4}}.$$

Introduce the random event  $\mathcal{A} = \left\{ \hat{d}_\varepsilon(\hat{\vartheta}_\varepsilon) \leq (\lambda \varphi_\varepsilon(\alpha_\varepsilon))^2 \right\}$ , and put

$$\hat{\varrho}_\varepsilon = \begin{cases} \varphi_\varepsilon(\alpha_\varepsilon) & \mathcal{A}_\varepsilon \text{ holds,} \\ \varphi_\varepsilon(\beta) & \mathcal{A}_\varepsilon^c \text{ holds;} \end{cases}$$

$$f_\varepsilon^*(\cdot) = \begin{cases} \hat{f}_\varepsilon^{(0)}(\cdot) & \mathcal{A}_\varepsilon \text{ holds,} \\ \hat{f}_\varepsilon(\cdot, \beta, Q) & \mathcal{A}_\varepsilon^c \text{ holds,} \end{cases}$$

where  $\mathcal{A}_\varepsilon^c$  is the event, being complement to the event  $\mathcal{A}_\varepsilon$ , and  $\hat{f}_\varepsilon^{(0)}(\cdot) = \bar{f}_\varepsilon(\cdot, \hat{\vartheta}_\varepsilon)$ .

For  $\forall z \in (0, \infty)$ ,  $\forall u \in (0, 1)$  and  $\forall \alpha \in [0, 1)$  define the function  $G_\alpha(u, z)$  as follows.

$$G_\alpha(u, z) = \sum_{k=0}^{\infty} \left( \frac{(4(m+1))^{\frac{1}{4}}}{z} \left( 1 + \sqrt{\frac{(k+1) \ln \frac{1}{u}}{\ln \frac{1}{\alpha}}} \right)^{\frac{1}{2}} + \frac{Q}{m!} z^{4\beta} \right)^q u^k.$$

If  $\alpha = 0$  we understand  $(\ln \frac{1}{\alpha})^{-1}$  as zero. Put

$$G_\alpha = \inf_{u \in (0, 1)} \inf_{z \in (0, \infty)} G_\alpha(u, z).$$

It is evident, that  $G_\alpha(\cdot, \cdot)$  is the smooth function and for  $\forall \alpha \neq 0$ ,  $G_\alpha(u, z) \rightarrow +\infty$ , if  $u \rightarrow 1$ ,  $u \rightarrow 0$ ,  $z \rightarrow 0$ ,  $z \rightarrow \infty$ . Therefore, there exist  $0 < Z_0 < \infty$  and  $0 < u_0 < 1$  such that

$$G_\alpha = G_\alpha(u_0, Z_0).$$

If  $Z_0$  is not unique, we take any  $Z_0$ , satisfying the last expression. If  $\alpha = 0$ , then  $u_0 = 0$  and  $Z_0$  can be calculated explicitly (see remark after Theorem 1). Remind, that  $Z_0$  is included in the definition of the threshold  $\lambda$ , describing the random event  $\mathcal{A}_\varepsilon$ , and in the definition of the variable  $N_\varepsilon$ , describing the estimator  $\hat{f}_\varepsilon(\cdot)$ .

**Theorem 1.** *Let assumptions **A1-A4** hold and let  $2^{-5} \geq \alpha_\varepsilon \geq \varepsilon^a$  for some fixed  $a > 0$  and for  $\forall \varepsilon \in (0, 1)$ , and let  $q \geq 2$ .*

*Then  $\hat{\varrho}_\varepsilon$  is optimal r.n.f. and  $f_\varepsilon^*(\cdot)$  is  $\alpha_\varepsilon$ -adaptive estimator. In particular, putting  $\liminf_{\varepsilon \rightarrow 0} \alpha_\varepsilon = \alpha \geq 0$ , we have for  $\forall q > 0$*

$$\limsup_{\varepsilon \rightarrow 0} R_\varepsilon^{(r)}(f_\varepsilon^*, \Sigma, \hat{\varrho}_\varepsilon) \leq M^*, \quad (5.10)$$

where

$$M^* = \left[ \frac{2\beta + 1}{(2\beta)^{\frac{2\beta}{2\beta+1}}} \left( \frac{(m+1)^\beta Q}{m!} \right)^{\frac{1}{2\beta+1}} \right]^q + G_\alpha.$$



**Remark 5.** First, let us notice, that if  $\alpha = 0$ , for example,  $\alpha_\varepsilon \rightarrow 0$  as  $\varepsilon \rightarrow 0$ , then  $Z_0$  can be calculated explicitly. Indeed, by definition,

$$G_0(u, z) = \left( \frac{(4(m+1))^{\frac{1}{4}}}{z} + \frac{Q}{m!} z^{4\beta} \right)^q \sum_{k=0}^{\infty} u^k = \left( \frac{(4(m+1))^{\frac{1}{4}}}{z} + \frac{Q}{m!} z^{4\beta} \right)^q \frac{1}{1-u}.$$

From here

$$Z_0 = \left( \frac{(4(m+1))^{\frac{1}{4}} m!}{4\beta Q} \right)^{\frac{1}{4\beta+1}},$$

$$G_0 = \left[ \frac{4\beta+1}{(4\beta)^{\frac{4\beta}{4\beta+1}}} \left( \frac{(4(m+1))^{\beta} Q}{m!} \right)^{\frac{1}{4\beta+1}} \right]^q,$$

and constant  $M^*$  has nice "symmetric" expression:

$$M^* = \left[ \frac{2\beta+1}{(2\beta)^{\frac{2\beta}{2\beta+1}}} \left( \frac{(m+1)^{\beta} Q}{m!} \right)^{\frac{1}{2\beta+1}} \right]^q + \left[ \frac{4\beta+1}{(4\beta)^{\frac{4\beta}{4\beta+1}}} \left( \frac{(4(m+1))^{\beta} Q}{m!} \right)^{\frac{1}{4\beta+1}} \right]^q.$$

Next, if  $q = 1$ , then expression for  $G_\alpha$  and for  $Z_0$  can be simplified.

$$Z_0 = \left( \frac{(4(m+1))^{\frac{1}{4}} m! \tilde{G}_\alpha}{4\beta Q} \right),$$

$$G_\alpha = \frac{4\beta+1}{(4\beta)^{\frac{4\beta}{4\beta+1}}} \left( \frac{(4(m+1))^{\beta} Q}{m!} \right)^{\frac{1}{4\beta+1}} (\tilde{G}_\alpha)^{\frac{4\beta}{4\beta+1}},$$

where

$$\tilde{G}_\alpha = \inf_{u \in (0,1)} \sum_{k=0}^{\infty} \left( 1 + \sqrt{\frac{(k+1) \ln \frac{1}{u}}{\ln \frac{1}{\alpha}}} \right)^{\frac{1}{2}} u^k.$$

**Corollary 1.** Note that all assumptions of Proposition 3 fulfill. Therefore, if **A1-A3** hold, using the decision rule, based on the event  $\mathcal{A}_\varepsilon$ , namely  $I\{\mathcal{A}_\varepsilon\}$ , we are able to test hypothesis  $H_0$  against the following set of local alternatives

$$\mathbf{H}_\varepsilon : f(\cdot) \in \Sigma : \inf_{\vartheta \in \Theta} \|f_0(\cdot, \vartheta) - f(\cdot)\|_2 \geq H \left( \varepsilon^4 \ln \frac{1}{\alpha_\varepsilon} \right)^{\frac{\beta}{4\beta+1}},$$

where  $H > 0$  is some given constant.

**Remark 6.** If  $\alpha_\varepsilon = \alpha \in (0,1)$  for  $\forall \varepsilon > 0$ , then the impossibility to construct more sensitive test follows from general lower bounds, obtained by Ingster (1993). In particular, it follows from the lower bound, proved in Theorem 1 for the risk  $R_\varepsilon^{(r)}(\cdot, \cdot, \cdot)$  under assumption **A4**. In the case  $\alpha_\varepsilon \rightarrow 0$ ,  $\varepsilon \rightarrow 0$ , this lower bound is new in the theory of hypothesis testing.

Let us briefly discuss the construction of the optimal r.n.f.  $\hat{\rho}_\varepsilon$  and  $\alpha(\cdot)$ -adaptive estimator  $f_\varepsilon^*$ . For each function  $f(\cdot) \in \Sigma$  let us define the following functional

$$d(f) = \inf_{\vartheta \in \Theta} \|f(\cdot) - f_0(\cdot, \vartheta)\|_2.$$

As it will follow from the further considerations,  $|\hat{d}_\varepsilon(\hat{\vartheta}_\varepsilon)|^{\frac{1}{2}}$  is the estimator of  $d(f)$ , providing  $\varphi_\varepsilon(\alpha_\varepsilon)$  as the accuracy of estimation on the space  $\Sigma$ . If  $\alpha_\varepsilon = \alpha > 0$  (or  $\inf_{f \in \Sigma_0} \alpha_\varepsilon > 0$ ), then  $\varphi_\varepsilon(\alpha_\varepsilon)$  coincides with MRC of functional  $d(f)$  on space  $\Sigma$ , and  $|\hat{d}_\varepsilon(\hat{\vartheta}_\varepsilon)|^{\frac{1}{2}}$  is the rate optimal estimator. If  $\alpha_\varepsilon \rightarrow 0$ ,  $\varepsilon \rightarrow 0$ , then  $\varphi_\varepsilon(\alpha_\varepsilon)$  differs from MRC of  $d(f)$  by the factor  $\left(\ln \frac{1}{\alpha_\varepsilon}\right)^{\frac{\beta}{4\beta+1}}$ . In some sense this factor is a payment for the following property of r.n.f.  $\hat{\vartheta}_\varepsilon$ :

$$\inf_{f(\cdot) \in \Sigma_0} \mathbf{P}_f \{ \hat{\vartheta}_\varepsilon = \varphi_\varepsilon(\alpha_\varepsilon) \} \geq 1 - \alpha_\varepsilon \rightarrow 1, \quad \varepsilon \rightarrow 0.$$

Remind, that by definition of optimal r.n.f. this payment is unavoidable. It is also evident, that  $d(f) = 0$  for  $\forall f(\cdot) \in \Sigma_0$ . So, if  $\hat{d}_\varepsilon(\hat{\vartheta}_\varepsilon)$  exceeds some threshold, it means, that  $d(f)$  is also not small, in other words, the hypothesis  $H_0$  is not true. In this case usual nonparametric estimator  $\hat{f}(\cdot, \beta, Q)$  is applied, and one can guarantee only MRC  $\varphi_\varepsilon(\beta)$ . If  $\hat{d}_\varepsilon(\hat{\vartheta}_\varepsilon)$  is less than  $\varphi_\varepsilon(\alpha_\varepsilon)$ , hence, estimated function lies not far from the set  $\Sigma_0$  as well. In this case, it makes sense to apply estimator  $\hat{f}_\varepsilon^{(0)}(\cdot)$ , which has nice behavior, if estimated function belongs to some neighborhood of the set  $\Sigma_0$ . In order to explain last words, let us consider the simple example, in which the estimator  $\hat{f}_\varepsilon^{(0)}(\cdot)$  can be found explicitly.

Suppose, that  $\Theta = \mathbf{R}^1$  and  $f_0(t, \vartheta) \equiv \vartheta, \forall t \in [0, 1]$ . It is easily to see that  $\bar{f}_\varepsilon(\cdot, \vartheta) \equiv \vartheta$  and, therefore  $\hat{f}_\varepsilon^{(0)}(\cdot) = \hat{\vartheta}_\varepsilon$ . In this case the minimization problem (5.9) has explicit solution

$$\hat{\vartheta}_\varepsilon = \left( \mathbf{v}' \Lambda^{-1} \mathbf{v} \right)^{-1} \frac{1}{N_\varepsilon} \sum_{k=1}^{N_\varepsilon} \left( \mathbf{x}^k \right)' \Lambda^{-1} \mathbf{v},$$

where  $\mathbf{v}' = (1, \frac{1}{2}, \dots, (m+1)^{-1})$ . Note that the vector  $\mathbf{v}$  is the first vector-column of the matrix  $\Lambda$ . Hence,  $\Lambda^{-1} \mathbf{v} = \mathbf{e}_1$ , where  $\mathbf{e}_1 = (1, 0, \dots, 0)$ . From here  $\mathbf{v}' \Lambda^{-1} \mathbf{v} = 1$ ;  $\left( \mathbf{x}^k \right)' \Lambda^{-1} \mathbf{v} = x_1^{(k)}$ . Hence,

$$\hat{\vartheta}_\varepsilon = \frac{1}{N_\varepsilon} \sum_{k=1}^{N_\varepsilon} x_1^{(k)} = \sum_{k=1}^{N_\varepsilon} \int_{\Delta_k} \mathbf{d}X_\varepsilon(t) = X_\varepsilon(1) - X_\varepsilon(0).$$

Thus,  $\hat{\vartheta}_\varepsilon = X_\varepsilon(1) - X_\varepsilon(0)$ . If the hypothesis  $H_0$  is really true, then  $\hat{\vartheta}_\varepsilon$  is  $\varepsilon$ -consistent, asymptotically efficient estimator. Now let us look at how  $\hat{\vartheta}_\varepsilon$  behaves "near" hypothesis set  $\Sigma_0$ .

Let  $U_{\varepsilon, \alpha_\varepsilon}(C) = \{f(\cdot) \in \Sigma : d(f) \leq C\varphi_\varepsilon(\alpha_\varepsilon)\}$ , where  $C > 0$  is some fixed constant. Note that in this case  $d(f) = \|f(\cdot) - \int_0^1 f(s) \mathbf{d}s\|_2$ . Therefore, for each  $f(\cdot) \in U_{\varepsilon, \alpha_\varepsilon}(C)$

$$\|\hat{\vartheta}_\varepsilon - f(\cdot)\|_2 = \left\| \int_0^1 f(s) \mathbf{d}s - f(\cdot) + \varepsilon(w(1) - w(0)) \right\|_2 \leq C\varphi_\varepsilon(\alpha_\varepsilon) + \varepsilon|w(1) - w(0)|.$$

In particular, from here

$$\sup_{f(\cdot) \in U_{\varepsilon, \alpha_\varepsilon}(C)} \mathbf{E}_f \left( \varphi_\varepsilon^{-1}(\alpha_\varepsilon) \|\hat{\vartheta}_\varepsilon - f(\cdot)\|_2 \right)^q \leq \left( C + \frac{\varepsilon}{\varphi_\varepsilon(\alpha_\varepsilon)} \sqrt{\frac{2}{\pi}} \right)^q \rightarrow C^q, \quad \varepsilon \rightarrow 0;$$

$$\sup_{f(\cdot) \in U_{\varepsilon, \alpha_\varepsilon}(C)} \mathbf{P}_f \left\{ \varphi_\varepsilon^{-1}(\alpha_\varepsilon) \|\hat{\vartheta}_\varepsilon - f(\cdot)\|_2 \geq C + \delta \right\} \asymp \exp \left\{ \left( -\frac{\varphi_\varepsilon(\alpha_\varepsilon)}{\varepsilon} \right)^2 \frac{\delta^2}{2} \right\},$$

$\forall \delta > 0$ , as  $\varepsilon \rightarrow 0$ . Last two lines show optimal, in some sense, behavior of the estimator  $\hat{\vartheta}_\varepsilon$  "near" hypothesis set  $\Sigma_0$ .

**5.2. Hypothesis on smoothness.** In this section we consider the hypothesis, described in the example 4.1.2.

$$\mathbf{H}_0 : f(\cdot) \in \Sigma_0 = H(\gamma, P) \cap \Sigma, \quad \beta < \gamma < \infty, \quad P > 0,$$

where  $H(\gamma, P)$  is Holder space. Thus, the hypothesis set  $\Sigma_0$  consists of the functions, which are smoother than functions, belonging to the set  $\Sigma$ . Here the constants  $\gamma = l + \nu$  ( $l \geq 0$  is integer,  $0 < \nu \leq 1$ ) and  $P$  are supposed to be known.

Let us fix some function  $0 < \alpha_\varepsilon < 1 - \delta$ ,  $0 < \delta < 1$ ,  $\varepsilon \in (0, 1)$ , and let again  $\Omega_\varepsilon$  be the family of r.n.f.  $\varrho_\varepsilon$ , satisfying the inequality

$$\limsup_{\varepsilon \rightarrow 0} \sup_{f(\cdot) \in \Sigma_0} \alpha_\varepsilon^{-1} \mathbf{P}_f \{ \varrho_\varepsilon = \varphi_\varepsilon(\beta) \} \leq 1. \quad (5.11)$$

We keep some notations from the previous section. In particular, for  $\forall \varrho_\varepsilon \in \Omega_\varepsilon$  and for an arbitrary estimator  $\tilde{f}_\varepsilon(\cdot)$ ,  $R_\varepsilon^{(r)}(\tilde{f}_\varepsilon, \Sigma, \varrho_\varepsilon)$  is the minimax risk, defined by formulae (5.3),  $\varphi_\varepsilon(\beta) = \varepsilon^{\frac{2\beta}{2\beta+1}}$ . Also, for  $\forall N \geq 1$ ,  $t_k = t_k(N) = k/N$ ,  $\Delta_k = [t_{k-1}, t_k)$ ,  $f_N(\cdot)$  is defined by (5.4) and (5.5).  $N_\varepsilon(\beta, Q)$  and  $\hat{f}_\varepsilon(\cdot, \beta, Q)$  remain the same as before.

Put for  $\forall z \in (0, \infty)$ ,  $\forall u \in (0, \infty)$  and  $\forall \alpha \in [0, 1)$

$$G_\alpha(z) = \sup_{u \geq 0} \left( \frac{(4(l+1))^{1/4}}{z} \left( 1 + \sqrt{\frac{u}{\ln \frac{1}{\alpha}}} \right)^{1/2} + \frac{Q}{m!} z^{4\beta} \right)^q \exp\{-u\};$$

$$Z_1 = \operatorname{arginf}_{z > 0} G_\alpha(z).$$

It is obvious, that such  $0 < Z_1 < \infty$  exist. (If  $Z_1$  is not unique, let us take any  $Z_1$ , satisfying the last equation.) Denote

$$N_\varepsilon = \left\lceil Z_1^{-4} \left( \varepsilon^4 \ln \frac{1}{\alpha_\varepsilon} \right)^{-\frac{1}{4\beta+1}} \right\rceil;$$

$$\hat{f}_\varepsilon(t) = f_{N_\varepsilon}(t) = \sum_{k=1}^{N_\varepsilon} (\mathbf{y}^{(k)})' \Lambda_l^{-\frac{1}{2}} \mathbf{d}^{(k)}(t) I\{t \in \Delta_k\},$$

where the vectors  $\mathbf{y}^{(k)}$ ,  $k = \overline{1, N_\varepsilon}$ , are the same as in previous section, and  $\Lambda_l$  is the  $(l+1) \times (l+1)$ -matrix with elements  $\lambda_{i,j} = (i+j-1)^{-1}$ ,  $i, j = \overline{1, l+1}$ . Thus, in comparison with the definition of the estimator  $\tilde{f}_\varepsilon(\cdot)$  in the previous section only the following changes were made: we replaced the constant  $Z_0$  by the constant

$Z_1$  for the definition of  $N_\varepsilon$  and replaced the matrix  $\Lambda_m$  by the matrix  $\Lambda_l$  in the formulae (5.5). Put

$$K_\varepsilon = Z_1^{-4} \left( \frac{\sqrt{l+1}!}{2P\gamma} \right)^{\frac{2}{2\gamma+1}} \left( \ln \frac{1}{\alpha_\varepsilon} \right)^{-\frac{1}{4\beta+1}} \varepsilon^{\frac{8(\gamma-\beta)+2}{(2\gamma+1)(4\beta+1)}}.$$

Without loss of generality  $K_\varepsilon$  is assumed to be integer. For  $\forall k = 1, \dots, \tilde{N}_\varepsilon$  put  $\tilde{t}_k = t_{kK_\varepsilon}$  and  $\tilde{\Delta}_k = [\tilde{t}_{k-1}, \tilde{t}_k)$ , where

$$\tilde{N}_\varepsilon = \left[ \frac{N_\varepsilon}{K_\varepsilon} \right] = \left[ \left( \frac{2P\gamma}{\sqrt{l+1}!} \right)^{\frac{2}{2\gamma+1}} \varepsilon^{-\frac{2}{2\gamma+1}} \right].$$

Introduce the vectors  $(\tilde{\mathbf{x}}^{(k)})' = (\tilde{x}_1^{(k)}, \dots, \tilde{x}_{l+1}^{(k)})$  and  $(\tilde{\mathbf{y}}^{(k)})' = (\tilde{y}_1^{(k)}, \dots, \tilde{y}_{l+1}^{(k)})$  as

$$\begin{aligned} \tilde{x}_j^{(k)} &= \tilde{N}_\varepsilon^j \int_{\tilde{\Delta}_k} (t - \tilde{t}_{k-1}) \mathbf{d}X_\varepsilon(t), \quad j = \overline{1, l+1}; \\ \tilde{\mathbf{y}}^{(k)} &= \Lambda_l^{-\frac{1}{2}} \tilde{\mathbf{x}}^{(k)}. \end{aligned}$$

Put also for  $\forall k = \overline{1, \tilde{N}_\varepsilon}$  and  $\forall t \in [0, 1]$ ,  $(\tilde{\mathbf{d}}^{(k)}(t))' = (1, \tilde{N}_\varepsilon(t - \tilde{t}_{k-1}), \dots, (\tilde{N}_\varepsilon(t - \tilde{t}_{k-1}))^l)$  and denote

$$\tilde{f}_\varepsilon(t) = \sum_{k=1}^{\tilde{N}_\varepsilon} (\tilde{\mathbf{y}}^{(k)})' \Lambda_l^{-\frac{1}{2}} \tilde{\mathbf{d}}^{(k)}(t) I\{t \in \tilde{\Delta}_k\}.$$

Note that  $\tilde{f}_\varepsilon(\cdot)$  is the estimator, which attains the MRC  $\varphi_\varepsilon(\gamma) = \varepsilon^{\frac{2\gamma}{2\gamma+1}}$  on the space  $H(\gamma, P)$ .

Now let us construct the optimal r.n.f.  $\hat{\varrho}_\varepsilon \in \Omega_\varepsilon$  and  $\alpha_{(\cdot)}$ -adaptive estimator. Put

$$\begin{aligned} \hat{d}_\varepsilon &= \|\hat{f}_\varepsilon(\cdot) - \tilde{f}_\varepsilon(\cdot)\|_2^2 + \varepsilon^2(l+1)(\tilde{N}_\varepsilon - N_\varepsilon); \\ \psi_\varepsilon(\alpha_\varepsilon) &= \left( \varepsilon^4 \ln \frac{1}{\alpha_\varepsilon} \right)^{\frac{\beta}{4\beta+1}}; \quad \varphi_\varepsilon(\gamma) = \varepsilon^{\frac{2\gamma}{2\gamma+1}}; \quad \tau_\varepsilon = \varepsilon^{\frac{4\gamma}{4\beta+1}}, \end{aligned}$$

and introduce the random event

$$\mathcal{A}_\varepsilon = \left\{ \hat{d}_\varepsilon \leq (\lambda \psi_\varepsilon(\alpha_\varepsilon))^2 + (\lambda_1 \varphi_\varepsilon(\gamma) + \lambda_2 \tau_\varepsilon)^2 \right\},$$

where

$$\lambda = \frac{(4(l+1))^{\frac{1}{4}}}{Z_1}; \quad \lambda_1 = \left( \frac{P}{l!} \right)^{\frac{1}{2\gamma+1}} \left( \frac{\sqrt{l+1}}{2\gamma} \right)^{\frac{2\gamma}{2\gamma+1}}; \quad \lambda_2 = \frac{P}{l!} Z_1^{4\gamma}.$$

Denote

$$\varphi_\varepsilon(\alpha_\varepsilon) = \begin{cases} \left( \varepsilon^4 \ln \frac{1}{\alpha_\varepsilon} \right)^{\frac{\beta}{4\beta+1}} & \gamma \geq 2\beta; \\ \varphi_\varepsilon(\gamma) & \beta < \gamma < 2\beta, \end{cases}$$

and put

$$\hat{\varrho}_\varepsilon = \begin{cases} \varphi_\varepsilon(\alpha_\varepsilon) & \mathcal{A}_\varepsilon \text{ holds;} \\ \varphi_\varepsilon(\beta) & \mathcal{A}_\varepsilon^c \text{ holds.} \end{cases}$$

$$f_\varepsilon^*(\cdot) = \begin{cases} \tilde{f}_\varepsilon(\cdot) & \mathcal{A}_\varepsilon \text{ holds;} \\ \hat{f}_\varepsilon(\cdot, \beta, Q) & \mathcal{A}_\varepsilon^c \text{ holds.} \end{cases}$$

**Theorem 2.** *Let  $q \geq 2$  and let  $2^{-5} \geq \alpha_\varepsilon \geq \varepsilon^a$ , for  $\forall \varepsilon > 0$ , for some  $a > 0$ .*

*Then  $\hat{\varrho}_\varepsilon$  is optimal r.n.f. and  $f_\varepsilon^*(\cdot)$  is  $\alpha_{(\cdot)}$ -adaptive estimator. In particular, putting  $\liminf_{\varepsilon \rightarrow 0} \alpha_\varepsilon = \alpha > 0$ , one can state for  $\forall q > 0$*

$$\limsup_{\varepsilon \rightarrow 0} R_\varepsilon^{(r)}(f_\varepsilon^*, \Sigma, \hat{\varrho}_\varepsilon) \leq M^*, \quad (5.12)$$

where for  $\beta < \gamma < 2\beta$

$$M^* = \left[ \frac{2\beta + 1}{(2\beta)^{\frac{2\beta}{2\beta+1}}} \left( \frac{(m+1)^\beta Q}{m!} \right)^{\frac{1}{2\beta+1}} \right]^q + \left[ \frac{2\gamma + 1}{(2\gamma)^{\frac{2\gamma}{2\gamma+1}}} \left( \frac{(l+1)^\gamma P}{l!} \right)^{\frac{1}{2\gamma+1}} \right]^q,$$

and for  $\gamma > 2\beta$

$$M^* = \left[ \frac{2\beta + 1}{(2\beta)^{\frac{2\beta}{2\beta+1}}} \left( \frac{(m+1)^\beta Q}{m!} \right)^{\frac{1}{2\beta+1}} \right]^q + G_\alpha(Z_1).$$

**Remark 7.** *For the case  $\gamma = 2\beta$  the constant  $M^*$  can be written explicitly as well, but the expression is too cumbersome, except the case  $\alpha = 0$ , see next remark.*

**Remark 8.** *If  $\alpha_\varepsilon \rightarrow 0$ , as  $\varepsilon \rightarrow 0$ , ( $\alpha = 0$ ) then the constants  $G_\alpha(z) = G_0(z)$ ,  $Z_1$  and, therefore,  $M^*$  can be calculated explicitly for the case  $\gamma \geq 2\beta$  as well.*

$$M^* = \left[ \frac{2\beta + 1}{(2\beta)^{\frac{2\beta}{2\beta+1}}} \left( \frac{(m+1)^\beta Q}{m!} \right)^{\frac{1}{2\beta+1}} \right]^q + \left[ \frac{4\beta + 1}{(4\beta)^{\frac{4\beta}{4\beta+1}}} \left( \frac{(4(l+1))^\beta Q}{m!} \right)^{\frac{1}{4\beta+1}} \right]^q,$$

**Corollary 2.** *Note, that all assumptions of Proposition 2 fulfill. Hence, if we choose  $\alpha_\varepsilon = \varepsilon^{\frac{2q\gamma}{2\gamma+1}}$ , then the estimator  $f_\varepsilon^*(\cdot)$  is the adaptive estimator.*

**Remark 9.** *As we see, for  $\beta < \gamma < 2\beta$  the function  $\varphi_\varepsilon(\alpha_{(\cdot)})$  coincides with MRC  $\varphi_\varepsilon(\gamma)$  and, due to Proposition 1, this is the best possible improvement of the accuracy of estimation. There is no any payment for the property  $\hat{\varrho}_\varepsilon \in \Omega_\varepsilon$ , since  $\varphi_\varepsilon(\alpha_\varepsilon) = \varphi_\varepsilon(\gamma)$  for  $\forall \alpha_{(\cdot)}$ , and it makes sense always to take  $\alpha_\varepsilon = \varepsilon^{\frac{2q\gamma}{2\gamma+1}}$  in view of Corollary 2. If  $2\beta < \gamma$  the improvement is worse than MRC, but one still exists, because  $\varphi_\varepsilon(\alpha_\varepsilon) = o(\varphi_\varepsilon(\beta))$ .*

## 6. Proofs of Theorem 1 and 2

**6.1. Proof of Theorem 1.** First, let us formulate several auxiliary lemmas. Put for  $\forall N \geq 1$ , for  $\forall k = \overline{1, N}$  and for  $\forall f(\cdot) \in \Sigma$

$$b_j^{(k)}(f) = N^j \int_{\Delta_k} (t - t_{k-1})^{j-1} f(t) dt; \quad j = \overline{1, m+1};$$

$$\left( \mathbf{b}^{(k)}(f) \right)' = \left( b_1^{(k)}(f), \dots, b_{m+1}^{(k)}(f) \right); \quad \mathbf{c}^{(k)}(f) = \Lambda^{-\frac{1}{2}} \mathbf{b}^{(k)}(f).$$

Denote for  $\forall t \in [0, 1]$

$$b_N(t, f) = \sum_{k=1}^N \left( \mathbf{b}^{(k)}(f) \right)' \Lambda^{-1} \mathbf{d}^{(k)}(t) I\{t \in \Delta_k\} \quad (6.1)$$

or in equivalent form

$$b_N(t, f) = \sum_{k=1}^N \left( \mathbf{c}^{(k)}(f) \right)' \Lambda^{-\frac{1}{2}} \mathbf{d}^{(k)}(t) I\{t \in \Delta_k\} \quad (6.2)$$

**Lemma 1.** For  $\forall N \geq 1$  and  $\forall f(\cdot) \in \Sigma$

$$\|b_N(\cdot, f) - f(\cdot)\|_2 \leq \frac{Q}{m!} N^{-\beta}.$$

The lemma deals with the upper bound for the bias of the estimator, obtained by the method of piecewise polynomials. Such type of upper bounds are well-known, see, f.e., Korostelev and Tsybakov (1993), but for further purposes it is convenient for us to give very short, simple and self-contained proof of Lemma 1.

Let for  $\forall N \geq 1$ , for  $\forall k = \overline{1, N}$  and  $\forall j = \overline{1, m+1}$ ,  $u_j^{(k)}$  be i.i.d. random variables and  $u_1^{(1)} \sim \mathcal{N}(0, 1)$ . Put

$$S_N = \sum_{k=1}^N \sum_{j=1}^{m+1} \left( (u_j^{(k)})^2 - 1 \right).$$

**Lemma 2.** For  $\forall N \leq 1$ , for  $\forall C_N \in (0, \infty)$  and for  $\forall x \in [0, NC_N(m+1)]$

$$\begin{aligned} \mathbf{P}\{S_N \geq x\} &\leq \exp \left\{ -\frac{x^2}{4(m+1)N(1+C_N)} \right\}; \\ \mathbf{P}\{S_N \leq -x\} &\leq \exp \left\{ -\frac{x^2}{4(m+1)N(1+C_N)} \right\}. \end{aligned}$$

If  $C_N = o(N^{-\frac{1}{3}})$ ,  $N \rightarrow \infty$ , the statement of the lemma remains true, if we replace right sides of the both inequalities by

$$\exp \left\{ -\frac{x^2}{4(m+1)N} \right\} (1 + o(1)),$$

where  $o(1) \rightarrow 0$ ,  $N \rightarrow \infty$ , uniformly w.r.t  $x \in [0, NC_N(m+1)]$ .

**Lemma 3.** For  $\forall \delta > 0 \exists v_1 > 0, v_2 > 0$  such that for all small enough  $\varepsilon > 0$  the following inequality holds.

$$\sup_{f(\cdot) \in H(\beta, Q)} \mathbf{P}_f \left\{ \varphi_\varepsilon(\beta)^{-1} \|\hat{f}_\varepsilon(\cdot, \beta, Q) - f(\cdot)\|_2 \geq C + \delta \right\} \leq v_1 \exp \left\{ -v_2 \varepsilon^{-\frac{2}{2\beta+1}} \right\},$$

where

$$C = \frac{2\beta + 1}{(2\beta)^{\frac{2\beta}{2\beta+1}}} \left( \frac{(m+1)^\beta Q}{m!} \right)^{\frac{1}{2\beta+1}}.$$

Such kind of exponential inequalities for the distributions of the normalized  $L_p$ -losses of linear nonparametric estimators were found in Lepski (1991) for an arbitrary  $2 \leq p < \infty$ . Here we only precise the constant  $C$  for the case  $p = 2$  and so, the proof of Lemma 3 can be omitted.

For  $\forall k = \overline{1, N_\varepsilon}$  introduce the vectors  $(\boldsymbol{\xi}^{(k)}(\varepsilon))' = (\xi_1^{(k)}(\varepsilon), \dots, \xi_{m+1}^{(k)}(\varepsilon))$  and  $(\boldsymbol{\eta}^{(k)}(\varepsilon))' = (\eta_1^{(k)}(\varepsilon), \dots, \eta_{m+1}^{(k)}(\varepsilon))$  as follows.

$$\xi_j^{(k)}(\varepsilon) = N^{j-\frac{1}{2}} \int_{\Delta_k} (t - t_{k-1}) \mathbf{d}w(t), \quad j = \overline{1, m+1} \quad (6.3)$$

$$\boldsymbol{\eta}^{(k)}(\varepsilon) = \Lambda^{-\frac{1}{2}} \boldsymbol{\xi}^{(k)}(\varepsilon). \quad (6.4)$$

Denote for  $\forall f(\cdot) \in \Sigma$

$$\begin{aligned} \Delta_\varepsilon^2(f) &= \frac{1}{N_\varepsilon} \sum_{k=1}^{N_\varepsilon} \|\mathbf{c}^{(k)}(f) - \mathbf{y}^{(k)}(\hat{\vartheta}_\varepsilon)\|^2; \\ z_\varepsilon(f) &= \frac{2\varepsilon}{\sqrt{N_\varepsilon}} \sum_{k=1}^{N_\varepsilon} (\mathbf{c}^{(k)}(f) - \mathbf{y}^{(k)}(\hat{\vartheta}_\varepsilon))' \boldsymbol{\eta}^{(k)}(\varepsilon). \end{aligned}$$

**Lemma 4.** Fix arbitrary  $0 < \delta_1 < \frac{1}{2}$  and  $\delta_2 > 0$ . There exist constants  $v_3 > 0$ ,  $v_4 > 0$  and  $\mu > 0$  such that

$$\sup_{f(\cdot) \in \Sigma} \mathbf{P} \left\{ |z_\varepsilon(f)| > \delta_1 \Delta_\varepsilon^2(f); \Delta_\varepsilon^2(f) \geq (\delta_2 \varphi_\varepsilon(\alpha_\varepsilon))^2 \right\} \leq v_3 \exp\{-v_4 \varepsilon^{-\mu}\},$$

for all small enough  $\varepsilon > 0$ .

6.1.1. *Proof of Theorem 1.* We divide the proof of the theorem on two steps. First, we prove upper bound, i.e. inequality (1.8) of Definition 1. To do this, it is enough to show that  $\hat{\varrho}_\varepsilon \in \Omega_\varepsilon$  and to obtain the inequality (5.10). Next, we find the lower bound, i.e. prove that r.n.f.  $\hat{\varrho}_\varepsilon$  can not be improved in the sense of the statement (1.9) of Definition 1.

**I(Upper bound).** 1. Let us prove, that  $\hat{\varrho}_\varepsilon \in \Omega_\varepsilon$ . By definition, it is required to show that

$$\limsup_{\varepsilon \rightarrow 0} \alpha_\varepsilon^{-1} \sup_{\vartheta \in \Theta} \mathbf{P}_\vartheta \{ \varrho_\varepsilon = \varphi_\varepsilon(\beta) \} \leq 1. \quad (6.5)$$

It is obvious, that  $\forall \vartheta \in \Theta$

$$\begin{aligned} \mathbf{P}_\vartheta \{ \hat{\varrho}_\varepsilon = \varphi_\varepsilon(\beta) \} &= \mathbf{P}_\vartheta \{ \hat{d}_\varepsilon(\hat{\vartheta}_\varepsilon) \geq (\lambda \varphi_\varepsilon(\alpha_\varepsilon))^2 \} \\ &= \mathbf{P}_\vartheta \left\{ \inf_{\vartheta \in \Theta} \hat{d}_\varepsilon(\vartheta) \geq (\lambda \varphi_\varepsilon(\alpha_\varepsilon))^2 \right\} \leq \mathbf{P}_\vartheta \{ \hat{d}_\varepsilon(\vartheta) \geq (\lambda \varphi_\varepsilon(\alpha_\varepsilon))^2 \}. \end{aligned} \quad (6.6)$$

As it follows from (6.6), in order to prove (6.5), it is enough to show that

$$\limsup_{\varepsilon \rightarrow 0} \alpha_\varepsilon^{-1} \sup_{\vartheta \in \Theta} \mathbf{P}_\vartheta \left\{ \hat{d}_\varepsilon(\vartheta) \geq (\lambda \varphi_\varepsilon(\alpha_\varepsilon))^2 \right\} \leq 1. \quad (6.7)$$

Note, that for  $\forall \varepsilon > 0$ , providing  $N_\varepsilon \geq 1$ , the vectors  $\boldsymbol{\xi}^{(k)}(\varepsilon)$ ,  $k = \overline{1, N_\varepsilon}$ , defined in (6.3), are i.i.d. Gaussian random vectors and  $\mathbf{E}\boldsymbol{\xi}^{(k)}(\varepsilon) = 0$ ,  $\mathbf{E}\boldsymbol{\xi}^{(k)}(\varepsilon) (\boldsymbol{\xi}^{(k)}(\varepsilon))' = \Lambda$ . Therefore, for all small enough  $\varepsilon > 0$  the vectors  $\boldsymbol{\eta}^{(k)}(\varepsilon)$ ,  $k = \overline{1, N_\varepsilon}$ , defined in

(6.4), are i.i.d. Gaussian vectors and  $\mathbf{E}\boldsymbol{\eta}^{(k)}(\varepsilon) = 0$ ,  $\mathbf{E}\boldsymbol{\eta}^{(k)}(\varepsilon) \left(\boldsymbol{\eta}^{(k)}(\varepsilon)\right)' = I_m$ , where  $I_m$  is  $(m+1) \times (m+1)$  unit matrix. From this observation, for  $\forall k = \overline{1, N_\varepsilon}$  and  $\forall j = \overline{1, m+1}$ ,  $\eta_j^{(k)}(\varepsilon)$  are i.i.d random variables and

$$\eta_j^{(k)}(\varepsilon) \sim \mathcal{N}(0, 1). \quad (6.8)$$

For  $\forall \vartheta \in \Theta$

$$\hat{d}_\varepsilon(\vartheta) = d_\varepsilon^2(\vartheta) - \varepsilon^2 N_\varepsilon(m+1) = \left( \frac{1}{N_\varepsilon} \sum_{k=1}^{N_\varepsilon} \|\mathbf{y}^{(k)} - \mathbf{y}^{(k)}(\vartheta)\|^2 \right) - \varepsilon^2 N_\varepsilon(m+1). \quad (6.9)$$

Here we used (5.7). Note, that for  $\forall f(\cdot)$  and for  $\forall k = \overline{1, N_\varepsilon}$

$$\mathbf{y}^{(k)} = \mathbf{c}^{(k)}(f) + \varepsilon \sqrt{N_\varepsilon} \boldsymbol{\eta}^{(k)}(\varepsilon), \quad (6.10)$$

and, moreover, for  $\forall \vartheta \in \Theta$

$$\mathbf{c}^{(k)}(f_0(\cdot, \vartheta)) = \mathbf{y}^{(k)}(\vartheta). \quad (6.11)$$

Hence, from (6.9), (6.10) and (6.11), for  $\forall \vartheta \in \Theta$  one has

$$\begin{aligned} \mathbf{P}_\vartheta \left\{ \hat{d}_\varepsilon(\vartheta) \geq (\lambda \varphi_\varepsilon(\alpha_\varepsilon))^2 \right\} &= \mathbf{P}_\vartheta \left\{ \sum_{k=1}^{N_\varepsilon} \left( \|\boldsymbol{\eta}^{(k)}(\varepsilon)\|^2 - (m+1) \right) \geq \left( \frac{\lambda \varphi_\varepsilon(\alpha_\varepsilon)}{\varepsilon} \right)^2 \right\} \\ &= \mathbf{P}_\vartheta \left\{ \sum_{k=1}^{N_\varepsilon} \sum_{j=1}^{m+1} \left( (\eta_j^{(k)}(\varepsilon))^2 - 1 \right) \geq \left( \frac{\lambda \varphi_\varepsilon(\alpha_\varepsilon)}{\varepsilon} \right)^2 \right\} \\ &= \mathbf{P}_\vartheta \left\{ S(N_\varepsilon) \geq \left( \frac{\lambda \varphi_\varepsilon(\alpha_\varepsilon)}{\varepsilon} \right)^2 \right\}, \end{aligned} \quad (6.12)$$

where

$$S(N_\varepsilon) = \sum_{k=1}^{N_\varepsilon} \sum_{j=1}^{m+1} \left( (\eta_j^{(k)}(\varepsilon))^2 - 1 \right). \quad (6.13)$$

Due to (6.8), for each fixed  $\varepsilon > 0$  such that  $N_\varepsilon \geq 1$ , for  $\forall k = \overline{1, N_\varepsilon}$  and  $\forall j = \overline{1, m+1}$  the distribution of the random variable  $\eta_j^{(k)}(\varepsilon)$  does not depend on  $\varepsilon$  and  $\vartheta$ . Therefore, we can apply Lemma 2 to the sum  $S(N_\varepsilon)$ , where one needs to put  $x = x_\varepsilon = \left( \frac{\lambda \varphi_\varepsilon(\alpha_\varepsilon)}{\varepsilon} \right)^2$  and to choose  $C_N = C_{N_\varepsilon}$ . Since

$$x_\varepsilon = (4(m+1))^{\frac{1}{2}} N_\varepsilon^{\frac{1}{2}} \sqrt{\ln \frac{1}{\alpha_\varepsilon}} \leq (4(m+1))^{\frac{1}{2}} N_\varepsilon^{\frac{1}{2}} \sqrt{a \ln \frac{1}{\varepsilon}} \quad (6.14)$$

(here we used, that  $\alpha_\varepsilon \geq \varepsilon^a$  by the assumption of the theorem) we can choose  $C_\varepsilon$  on such a way, that  $C_\varepsilon = o(N_\varepsilon^{-\frac{1}{3}})$  as  $\varepsilon \rightarrow 0$ , and  $0 \leq x_\varepsilon \leq N_\varepsilon C_\varepsilon(m+1)$ , for  $\forall \varepsilon > 0$ , which is small enough. Thus, one can apply Lemma 2. It yields together with (6.12)

$$\mathbf{P}_\vartheta \left\{ \hat{d}_\varepsilon(\vartheta) \geq (\lambda \varphi_\varepsilon(\alpha_\varepsilon))^2 \right\} \leq \exp \left\{ - \left( \frac{\lambda \varphi_\varepsilon(\alpha_\varepsilon)}{\varepsilon} \right)^4 \frac{1}{4N_\varepsilon(m+1)} \right\} (1 + o(1)), \quad (6.15)$$



as,  $\varepsilon \rightarrow 0$ , uniformly w.r.t.  $\vartheta \in \Theta$ . Since

$$\left( \frac{\lambda \varphi_\varepsilon(\alpha_\varepsilon)}{\varepsilon} \right)^4 \frac{1}{4N_\varepsilon(m+1)} = \ln \frac{1}{\alpha_\varepsilon},$$

finally we get from (6.15)

$$\sup_{\vartheta \in \Theta} \mathbf{P}_\vartheta \left\{ \hat{d}_\varepsilon(\vartheta) \geq (\lambda \varphi_\varepsilon(\alpha_\varepsilon))^2 \right\} \leq \alpha_\varepsilon(1 + o(1)), \quad (6.16)$$

as  $\varepsilon \rightarrow 0$ . This completes the prof of (6.7) and, therefore, (6.5).

**2.** Now let us prove the inequality (5.10). Put for  $\forall f(\cdot) \in \Sigma$

$$\begin{aligned} R_\varepsilon^{(1)}(f) &= \mathbf{E}_f \left\{ \hat{\varrho}_\varepsilon^{-1} \|f_\varepsilon^*(\cdot) - f(\cdot)\|_2 \right\}^q I\{\mathcal{A}_\varepsilon\} \\ &= \mathbf{E}_f \left\{ \varphi_\varepsilon^{-1}(\alpha_\varepsilon) \|\hat{f}_\varepsilon^{(0)}(\cdot) - f(\cdot)\|_2 \right\}^q I\{\mathcal{A}_\varepsilon\}; \end{aligned} \quad (6.17)$$

$$\begin{aligned} R_\varepsilon^{(2)}(f) &= \mathbf{E}_f \left\{ \hat{\varrho}_\varepsilon^{-1} \|f_\varepsilon^*(\cdot) - f(\cdot)\|_2 \right\}^q I\{\mathcal{A}_\varepsilon^c\} \\ &= \mathbf{E}_f \left\{ \varphi_\varepsilon^{-1}(\beta) \|\hat{f}_\varepsilon(\cdot, \beta, Q) - f(\cdot)\|_2 \right\}^q I\{\mathcal{A}_\varepsilon^c\}. \end{aligned} \quad (6.18)$$

It is evident, that

$$R_\varepsilon^{(r)}(f_\varepsilon^*, \Sigma, \hat{\varrho}_\varepsilon) \leq \sup_{f(\cdot) \in \Sigma} R_\varepsilon^{(1)}(f) + \sup_{f(\cdot) \in \Sigma} R_\varepsilon^{(2)}(f), \quad (6.19)$$

and, hence, it is enough to obtain upper estimates for  $R_\varepsilon^{(1)}(\cdot)$  and for  $R_\varepsilon^{(2)}(\cdot)$ . First, let us estimate  $R_\varepsilon^{(2)}(\cdot)$ . Clearly, that

$$\begin{aligned} R_\varepsilon^{(2)} &= \sup_{f(\cdot) \in \Sigma} R_\varepsilon^{(2)}(f) \\ &\leq \sup_{f(\cdot) \in \Sigma} \mathbf{E}_f \left\{ \varphi_\varepsilon^{-1}(\beta) \|\hat{f}_\varepsilon(\cdot, \beta, Q) - f(\cdot)\|_2 \right\}^q. \end{aligned} \quad (6.20)$$

From Lemma 3 one immediately has

$$\sup_{f(\cdot) \in \Sigma} \mathbf{E}_f \left\{ \varphi_\varepsilon^{-1}(\beta) \|\hat{f}_\varepsilon(\cdot, \beta, Q) - f(\cdot)\|_2 \right\}^q \leq (C + \delta)^q (1 + o(1)), \quad (6.21)$$

as  $\varepsilon \rightarrow 0$  for  $\forall \delta > 0$ . From (6.20) and (6.21)

$$\limsup_{\varepsilon \rightarrow 0} R_\varepsilon^{(2)} \leq (C + \delta)^q.$$

Since  $\delta$  is an arbitrary number, tending  $\delta$  to zero, finally we get

$$\limsup_{\varepsilon \rightarrow 0} \sup_{f(\cdot) \in \Sigma} R_\varepsilon^{(2)}(f) \leq \left( \frac{2\beta + 1}{2\beta \frac{2\beta}{2\beta+1}} \left( \frac{(m+1)^\beta Q}{m!} \right)^{\frac{1}{2\beta+1}} \right)^q. \quad (6.22)$$

Let us now start with estimation of  $R_\varepsilon^{(1)}(\cdot)$ . Fix some  $0 < \delta_1 < \frac{1}{2}$ ,  $0 < u < 1$ , and put for  $\forall k = \overline{1, k(\varepsilon)}$ ,  $k(\varepsilon) = q \ln(\varepsilon) / \ln(u)$

$$T_k = \left( \sqrt{\frac{k \ln(u)}{\ln(\alpha_\varepsilon)}} + 1 \right) (\lambda \varphi_\varepsilon(\alpha_\varepsilon))^2.$$

Introduce the family of the random events  $\Gamma_k$ ,  $k = \overline{1, k(\varepsilon) - 1}$  as follows.

$$\begin{aligned}\Gamma_0 &= \{(1 - \delta_1)\Delta_\varepsilon^2(f) \leq T_1\}; \quad \Gamma_{k(\varepsilon)} = \{(1 - \delta_1)\Delta_\varepsilon^2(f) \geq T_{k(\varepsilon)}\}; \\ \Gamma_k &= \{T_k \leq (1 - \delta_1)\Delta_\varepsilon^2(f) \leq T_{k+1}\}.\end{aligned}$$

Note, that

$$\Gamma_0^c = \bigcup_{k=1}^{k(\varepsilon)} \Gamma_k; \quad \Gamma_k \cap \Gamma_l = \emptyset, \quad \forall k, l = \overline{0, k(\varepsilon)}, \quad k \neq l. \quad (6.23)$$

Denote also

$$D_\varepsilon = \{|z_\varepsilon(f)| \leq \delta_1 \Delta_\varepsilon^2(f)\}.$$

We have by the triangle inequality

$$\begin{aligned}\varphi_\varepsilon^{-1}(\alpha_\varepsilon) \|\hat{f}_\varepsilon^{(0)}(\cdot) - f(\cdot)\|_2 &= \varphi_\varepsilon^{-1}(\alpha_\varepsilon) \|\bar{f}_\varepsilon(\cdot, \hat{\vartheta}_\varepsilon) - f(\cdot)\|_2 \\ &\leq \varphi_\varepsilon^{-1}(\alpha_\varepsilon) \|\bar{f}_\varepsilon(\cdot, \hat{\vartheta}_\varepsilon) - b_{N_\varepsilon}(\cdot, f)\|_2 + \varphi_\varepsilon^{-1}(\alpha_\varepsilon) \|b_{N_\varepsilon}(\cdot, f) - f(\cdot)\|_2.\end{aligned} \quad (6.24)$$

Here  $b_{N_\varepsilon}(\cdot, f)$  is defined by formulae (6.1) or (6.2), where one needs to put  $N = N_\varepsilon$ . Then, from Lemma 1 for  $\forall f(\cdot) \in \Sigma$  one obtains

$$\varphi_\varepsilon^{-1}(\alpha_\varepsilon) \|b_{N_\varepsilon}(\cdot, f) - f(\cdot)\|_2 \leq A_1 = \frac{Q}{m!} \left(N_\varepsilon^\beta \varphi_\varepsilon(\alpha_\varepsilon)\right)^{-1} = A_1, \quad (6.25)$$

where  $A_1 = \frac{Q}{m!} Z_0^4$ . Note also that

$$\|b_{N_\varepsilon}(\cdot, f) - \bar{f}(\cdot, \hat{\vartheta}_\varepsilon)\|_2 = \Delta_\varepsilon^2(f), \quad (6.26)$$

and (6.23) together with (6.25) yields

$$\varphi_\varepsilon^{-1}(\alpha_\varepsilon) \|\hat{f}_\varepsilon^{(0)}(\cdot) - f(\cdot)\|_2 \leq \varphi_\varepsilon^{-1}(\alpha_\varepsilon) \Delta_\varepsilon(f) + A_1. \quad (6.27)$$

Put

$$\begin{aligned}R_\varepsilon^{(1,1)}(f) &= \mathbf{E}_f \left( \varphi_\varepsilon^{-1}(\alpha_\varepsilon) \Delta_\varepsilon(f) + A_1 \right)^q I\{\mathcal{A}_\varepsilon \cap \Gamma_0\}; \\ R_\varepsilon^{(1,2)}(f) &= \sum_{k=1}^{k(\varepsilon)-1} \mathbf{E}_f \left( \varphi_\varepsilon^{-1}(\alpha_\varepsilon) \Delta_\varepsilon(f) + A_1 \right)^q I\{\mathcal{A}_\varepsilon \cap \Gamma_k \cap D_\varepsilon\}; \\ R_\varepsilon^{(1,3)}(f) &= \mathbf{E}_f \left( \varphi_\varepsilon^{-1}(\alpha_\varepsilon) \Delta_\varepsilon(f) + A_1 \right)^q I\{\mathcal{A}_\varepsilon \cap \Gamma_{k(\varepsilon)} \cap D_\varepsilon\}; \\ R_\varepsilon^{(1,4)}(f) &= \mathbf{E}_f \left( \varphi_\varepsilon^{-1}(\alpha_\varepsilon) \Delta_\varepsilon(f) + A_1 \right)^q I\{\mathcal{A}_\varepsilon \cap \Gamma_0^c \cap D_\varepsilon^c\}.\end{aligned}$$

Due to (6.23)

$$\begin{aligned}I\{\mathcal{A}_\varepsilon\} &= I\{\mathcal{A}_\varepsilon \cap \Gamma_0\} + I\{\mathcal{A}_\varepsilon \cap \Gamma_0^c\} \\ &= I\{\mathcal{A}_\varepsilon \cap \Gamma_0\} + I\{\mathcal{A}_\varepsilon \cap \Gamma_0^c \cap D_\varepsilon^c\} + I\{\mathcal{A}_\varepsilon \cap \Gamma_0^c \cap D_\varepsilon\} \\ &= I\{\mathcal{A}_\varepsilon \cap \Gamma_0\} + I\{\mathcal{A}_\varepsilon \cap \Gamma_0^c \cap D_\varepsilon^c\} + I\{\mathcal{A}_\varepsilon \cap \Gamma_{k(\varepsilon)} \cap D_\varepsilon\} \\ &\quad + \sum_{k=1}^{k(\varepsilon)-1} I\{\mathcal{A}_\varepsilon \cap \Gamma_k \cap D_\varepsilon\}.\end{aligned}$$

From here and (6.27) for  $\forall f(\cdot) \in \Sigma$

$$R_\varepsilon^{(1)}(f) \leq \sum_{i=1}^4 R_\varepsilon^{(1,i)}(f). \quad (6.28)$$

Let us estimate  $R_\varepsilon^{(1,4)}(\cdot)$ . Remind that we consider the functions  $f(\cdot) \in \Sigma_1(L)$  and it is easily to check, there exists some constant  $L_1 = L_1(\beta, L)$  such that  $|\Delta_\varepsilon(f)| \leq L_1$  for  $\forall \varepsilon > 0$ . Hence,

$$\begin{aligned} R_\varepsilon^{(1,4)}(f) &\leq \left(L_1 \varphi_\varepsilon^{-1}(\alpha_\varepsilon) + A_1\right)^q \mathbf{P}_f \left\{ \mathcal{A}_\varepsilon \cap \Gamma_0^c \cap D_\varepsilon^c \right\} \\ &\leq \left(L_1 \varphi_\varepsilon^{-1}(\alpha_\varepsilon) + A_1\right)^q \mathbf{P}_f \left\{ \Gamma_0^c \cap D_\varepsilon^c \right\} \\ &= \left(L_1 \varphi_\varepsilon^{-1}(\alpha_\varepsilon) + A_1\right)^q \mathbf{P}_f \left\{ |z_\varepsilon(f)| \geq \delta_1 \Delta_\varepsilon^2(f); \Delta_\varepsilon^2(f) \geq \frac{T_1}{1 - \delta_1} \right\} \\ &= \left(L_1 \varphi_\varepsilon^{-1}(\alpha_\varepsilon) + A_1\right)^q \mathbf{P}_f \left\{ |z_\varepsilon(f)| \geq \delta_1 \Delta_\varepsilon^2(f); \Delta_\varepsilon^2(f) \geq \left( \frac{\lambda}{\sqrt{1 - \delta_1}} \varphi_\varepsilon(\alpha_\varepsilon) \right)^2 \right\}. \end{aligned}$$

Putting  $\delta_2 = \frac{\lambda}{\sqrt{1 - \delta_1}}$ , we have in view of Lemma 4

$$\sup_{f(\cdot) \in \Sigma} R_\varepsilon^{(1,4)}(f) \leq \left(L_1 \varphi_\varepsilon^{-1}(\alpha_\varepsilon) + A_1\right)^q v_3 \exp\{-v_4 \varepsilon^{-\mu}\},$$

and, therefore,

$$\limsup_{\varepsilon \rightarrow 0} \sup_{f(\cdot) \in \Sigma} R_\varepsilon^{(1,4)}(f) = 0. \quad (6.29)$$

Let us estimate  $R_\varepsilon^{(1,3)}(\cdot)$ . Quite analogously,

$$R_\varepsilon^{(1,3)}(f) \leq \left(L_1 \varphi_\varepsilon^{-1}(\alpha_\varepsilon) + A_1\right)^q \mathbf{P}_f \left\{ \mathcal{A}_\varepsilon \cap \Gamma_{k(\varepsilon)} \cap D_\varepsilon \right\}. \quad (6.30)$$

Note that by definition

$$\hat{d}(\hat{\vartheta}_\varepsilon) = \Delta_\varepsilon^2(f) - z_\varepsilon(f) + \varepsilon^2 S(N_\varepsilon),$$

where  $S(N_\varepsilon)$  is defined in (6.13). Therefore,

$$\begin{aligned} \left\{ \mathcal{A}_\varepsilon \cap \Gamma_{k(\varepsilon)} \cap D_\varepsilon \right\} &= \left\{ \hat{d}_\varepsilon(\hat{\vartheta}_\varepsilon) \leq (\lambda \varphi_\varepsilon(\alpha_\varepsilon))^2 \right\} \cap \left\{ \Gamma_{k(\varepsilon)} \right\} \cap \left\{ D_\varepsilon \right\} \\ &= \left\{ \Delta_\varepsilon^2(f) - z_\varepsilon(f) + \varepsilon^2 S(N_\varepsilon) \leq (\lambda \varphi_\varepsilon(\alpha_\varepsilon))^2 \right\} \\ &\cap \left\{ |z_\varepsilon(f)| \leq \delta_1 \Delta_\varepsilon^2(f) \right\} \cap \left\{ (1 - \delta_1) \Delta_\varepsilon^2(f) \geq T_{k(\varepsilon)} \right\} \\ &\subseteq \left\{ \varepsilon^2 S(N_\varepsilon) \leq -T_{k(\varepsilon)} + (\lambda \varphi_\varepsilon(\alpha_\varepsilon))^2 \right\} = \left\{ S(N_\varepsilon) \leq - \left( \frac{\lambda \varphi_\varepsilon(\alpha_\varepsilon)}{\varepsilon} \right)^2 \sqrt{\frac{q \ln \varepsilon}{\ln \alpha_\varepsilon}} \right\}. \end{aligned}$$

From here and (6.30) one has for  $\forall f(\cdot) \in \Sigma$

$$R_\varepsilon^{(1,3)}(f) \leq \left(L_1 \varphi_\varepsilon^{-1}(\alpha_\varepsilon) + A_1\right)^q \mathbf{P}_f \left\{ S(N_\varepsilon) \leq - \left( \frac{\lambda \varphi_\varepsilon(\alpha_\varepsilon)}{\varepsilon} \right)^2 \sqrt{\frac{q \ln \varepsilon}{\ln \alpha_\varepsilon}} \right\}. \quad (6.31)$$

As it has been already mentioned above, the distribution of the addends in  $S(N_\varepsilon)$  does not depend on  $f(\cdot)$  and  $\varepsilon$ . Hence, we can apply again Lemma 2 with

$$x = x_\varepsilon = \left( \frac{\lambda \varphi_\varepsilon(\alpha_\varepsilon)}{\varepsilon} \right)^2 \sqrt{\frac{q \ln \varepsilon}{\ln \alpha_\varepsilon}}$$

and properly chosen  $C_N = C_{N_\varepsilon}$ . It yields

$$\begin{aligned} & \mathbf{P}_f \left\{ S(N_\varepsilon) \leq - \left( \frac{\lambda \varphi_\varepsilon(\alpha_\varepsilon)}{\varepsilon} \right)^2 \sqrt{\frac{q \ln \varepsilon}{\ln \alpha_\varepsilon}} \right\} \\ & \leq \exp \left\{ - \left( \frac{\lambda \varphi_\varepsilon(\alpha_\varepsilon)}{\varepsilon} \right)^4 \frac{q \ln \varepsilon}{4N_\varepsilon(m+1) \ln \alpha_\varepsilon} \right\} (1 + o(1)) \\ & = \exp\{q \ln \varepsilon\} (1 + o(1)) = \varepsilon^q (1 + o(1)), \end{aligned}$$

as  $\varepsilon \rightarrow 0$ . Continuing (6.31), we arrive

$$\sup_{f(\cdot) \in \Sigma} R_\varepsilon^{(1,3)}(f) \leq (L_1 \varphi_\varepsilon^{-1}(\alpha_\varepsilon) + A_1)^q \varepsilon^q (1 + o(1)).$$

Since  $\varphi_\varepsilon(\alpha_\varepsilon) = \left( \varepsilon^4 \ln \frac{1}{\alpha_\varepsilon} \right)^{\frac{\beta}{4\beta+1}} \leq \left( \varepsilon^4 a \ln \frac{1}{\varepsilon} \right)^{\frac{\beta}{4\beta+1}}$ , finally we have

$$\limsup_{\varepsilon \rightarrow 0} \sup_{f(\cdot) \in \Sigma} R_\varepsilon^{(1,3)}(f) = 0. \quad (6.32)$$

Now, let us estimate  $R_\varepsilon^{(1,1)}(\cdot)$ . By definition

$$\begin{aligned} R_\varepsilon^{(1,1)}(f) &= \mathbf{E}_f \left( \varphi_\varepsilon^{-1}(\alpha_\varepsilon) \Delta_\varepsilon(f) + A_1 \right)^q I \left\{ \mathcal{A}_\varepsilon ; \Delta_\varepsilon^2(f) \leq \frac{T_1}{1 - \delta_1} \right\} \\ &\leq \left( \varphi_\varepsilon^{-1}(\alpha_\varepsilon) \sqrt{\frac{T_1}{1 - \delta_1}} + A_1 \right)^q = \left( \lambda \left( 1 + \sqrt{\frac{\ln u}{\ln \alpha_\varepsilon}} \right)^{\frac{1}{2}} \frac{1}{\sqrt{1 - \delta_1}} + A_1 \right)^q. \end{aligned}$$

From here, tending first  $\varepsilon$  to zero and then  $\delta_1$  to zero, one has

$$\limsup_{\varepsilon \rightarrow 0} \sup_{f(\cdot) \in \Sigma} R_\varepsilon^{(1,1)}(f) \leq \left( \lambda \left( 1 + \sqrt{\frac{\ln u}{\ln \alpha}} \right)^{\frac{1}{2}} + A_1 \right)^q, \quad (6.33)$$

for  $\forall u \in (0, 1)$ , where, remind,  $0 \leq \alpha = \liminf_{\varepsilon \rightarrow 0} \alpha_\varepsilon$ . At last, let us estimate  $R_\varepsilon^{(1,2)}(\cdot)$ . For  $\forall k = 1, k(\varepsilon) - 1$

$$\begin{aligned} & \{ \mathcal{A}_\varepsilon \cap \Gamma_k \cap D_\varepsilon \} \\ &= \left\{ \Delta_\varepsilon^2(f) - z_\varepsilon(f) + \varepsilon^2 S(N_\varepsilon) \leq (\lambda \varphi_\varepsilon(\alpha_\varepsilon))^2 \right\} \\ &\cap \left\{ T_k \leq (1 - \delta_1) \Delta_\varepsilon^2(f) \leq T_{k+1} \right\} \cap \left\{ |z_\varepsilon(f)| \leq \delta_1 \Delta_\varepsilon^2(f) \right\} \\ &\subseteq \left\{ \varepsilon^2 S(N_\varepsilon) \leq -T_k + (\lambda \varphi_\varepsilon(\alpha_\varepsilon))^2 \right\} \cap \left\{ (1 - \delta_1) \Delta_\varepsilon^2(f) \leq T_{k+1} \right\} \\ &= \left\{ S(N_\varepsilon) \leq - \left( \frac{\lambda \varphi_\varepsilon(\alpha_\varepsilon)}{\varepsilon} \right)^2 \sqrt{\frac{k \ln u}{\ln \alpha_\varepsilon}} \right\} \cap \left\{ (1 - \delta_1) \Delta_\varepsilon^2(f) \leq T_{k+1} \right\}. \end{aligned}$$

From here

$$\begin{aligned} R_\varepsilon^{(1,2)}(f) &\leq \sum_{k=1}^{k(\varepsilon)-1} \mathbf{E}_f \left( \varphi_\varepsilon^{-1}(\alpha_\varepsilon) \Delta_\varepsilon(f) + A_1 \right)^q I \left\{ B_\varepsilon^{(k)} ; (1 - \delta_1) \Delta_\varepsilon^2(f) \leq T_{k+1} \right\} \\ &\leq \sum_{k=1}^{k(\varepsilon)-1} \left( \varphi_\varepsilon^{-1}(\alpha_\varepsilon) \sqrt{\frac{T_{k+1}}{1 - \delta_1}} + A_1 \right)^q \mathbf{P}_f \left\{ B_\varepsilon^{(k)} \right\}, \end{aligned} \quad (6.34)$$

where  $B_\varepsilon^{(k)} = \left\{ S(N_\varepsilon) \leq - \left( \frac{\lambda \varphi_\varepsilon(\alpha_\varepsilon)}{\varepsilon} \right)^2 \sqrt{\frac{k \ln u}{\ln \alpha_\varepsilon}} \right\}$ . As before, Lemma 2 can be applied for  $\forall k = \overline{1, k(\varepsilon) - 1}$ . It yields

$$\begin{aligned} \mathbf{P}_f \{ B_\varepsilon^{(k)} \} &\leq \exp \left\{ - \left( \frac{\lambda \varphi_\varepsilon(\alpha_\varepsilon)}{\varepsilon} \right)^4 \frac{k \ln u}{4N_\varepsilon(m+1) \ln \alpha_\varepsilon} \right\} (1 + o(1)) \\ &= u^k (1 + o(1)), \end{aligned} \quad (6.35)$$

uniformly w.r.t  $k = \overline{1, k(\varepsilon) - 1}$ . We have from (6.34) and (6.35)

$$\begin{aligned} R_\varepsilon^{(1,2)}(f) &\leq \left( \sum_{k=1}^{k(\varepsilon)-1} \left( \lambda \left( 1 + \sqrt{\frac{(k+1) \ln u}{\ln \alpha_\varepsilon}} \right)^{\frac{1}{2}} \frac{1}{\sqrt{1 - \delta_1}} + A_1 \right)^q u^k \right) (1 + o(1)) \\ &\leq \left( \sum_{k=1}^{\infty} \left( \lambda \left( 1 + \sqrt{\frac{(k+1) \ln u}{\ln \alpha_\varepsilon}} \right)^{\frac{1}{2}} \frac{1}{\sqrt{1 - \delta_1}} + A_1 \right)^q u^k \right) (1 + o(1)) \end{aligned} \quad (6.36)$$

Since the series in the right side of (6.36) is absolutely converging for  $\forall u \in (0, 1)$  and for  $\forall \varepsilon > 0$ , we have

$$\limsup_{\varepsilon \rightarrow 0} \sup_{f(\cdot) \in \Sigma} R_\varepsilon^{(1,2)}(f) \leq \sum_{k=1}^{\infty} \left( \lambda \left( 1 + \sqrt{\frac{(k+1) \ln u}{\ln \alpha}} \right)^{\frac{1}{2}} + A_1 \right)^q u^k. \quad (6.37)$$

Here we took into account that  $\delta_1 > 0$  is an arbitrary number. Combining (6.29), (6.32), (6.33) and (6.37), from (6.28) we get

$$\limsup_{\varepsilon \rightarrow 0} \sup_{f(\cdot) \in \Sigma} R_\varepsilon^{(1)}(f) \leq G_\alpha(Z_0, u). \quad (6.38)$$

Taking into account, that left side in (6.38) does not depend on  $u \in (0, 1)$ , finally we have

$$\limsup_{\varepsilon \rightarrow 0} \sup_{f(\cdot) \in \Sigma} R_\varepsilon^{(1)}(f) \leq G_\alpha(Z_0, u_0) = G_\alpha. \quad (6.39)$$

Putting together (6.22) and (6.39), we arrive to the inequality (5.10).

**II(Lower bound).** Let  $\varrho_\varepsilon$  be an arbitrary r.n.f., taking two values  $\{\varphi_\varepsilon(\beta)\}$  and  $\{a_\varepsilon\}$ ,  $\varrho_\varepsilon \in \Omega_\varepsilon$ , and suppose that

$$\lim_{\varepsilon \rightarrow 0} \frac{a_\varepsilon}{\varphi_\varepsilon(\alpha_\varepsilon)} = 0. \quad (6.40)$$

In accordance with Definition 1 one needs to prove that for  $\forall \varrho_\varepsilon \in \Omega_\varepsilon$ , satisfying (6.40),

$$\liminf_{\varepsilon \rightarrow 0} \inf_{\tilde{f}_\varepsilon} R_\varepsilon^{(r)}(\tilde{f}_\varepsilon, \Sigma, \varrho_\varepsilon) = +\infty, \quad (6.41)$$

where infimum is taken over all possible estimators.

Let  $F(x), x \in \mathbf{R}^1$ , be the function, satisfying the following conditions

$$\begin{aligned} F(x) &= 0, \quad \forall x \in \mathbf{R}^1 \setminus [-1, 1]; \\ F(x) &\in H(\beta, 1), \quad x \in [-1, 1]; \\ \sigma^2 &= \int_{-1}^1 F^2(x) dx > 0. \end{aligned} \quad (6.42)$$

Put  $\delta_\varepsilon = \left( (\bar{Q}\sigma)^{-1} \varepsilon^4 \ln \frac{1}{\alpha_\varepsilon} \right)^{\frac{1}{4\beta+1}}$ ,  $k(\varepsilon) = 1/\delta_\varepsilon$  (without loss of generality  $k(\varepsilon)$  is assumed to be integer), where  $\bar{Q} = Q - Q_1$ ,  $Q_1$  is defined in the assumption **A4**. Let  $\vec{v}$  be the vector  $(v_1, \dots, v_{k(\varepsilon)})$  with coordinates, taking two values  $+1$  and  $-1$ . Denote by  $V$  the collection of vectors  $\vec{v}$ . Obvious,  $|V| = 2^{k(\varepsilon)}$ . Consider the family of functions  $\bar{f}(\cdot, \vec{v})$ ,  $\vec{v} \in V$ , defined on the interval  $[0, 1]$ :

$$\bar{f}(t, \vec{v}) = \sum_{i=1}^{k(\varepsilon)} v_i F_i(t),$$

where  $F_i(t) = \bar{Q} \delta_\varepsilon^\beta F\left(\frac{t-t_i}{\delta_\varepsilon}\right)$ ,  $t_i = i\delta_\varepsilon$ ,  $i = \overline{1, k(\varepsilon)}$ . Set also  $f_0(\cdot) = f_0(\cdot, \vartheta_0)$ ;  $f(\cdot, \vec{v}) = f_0(\cdot, \vartheta_0) + \bar{f}(\cdot, \vec{v})$ ,  $\vec{v} \in V$ , and introduce the following notation:

$$\begin{aligned} \mathbf{P}_0 &= \mathbf{P}_{f_0}; \quad \mathbf{P}_{\vec{v}} = \mathbf{P}_{f(\cdot, \vec{v})}, \quad \vec{v} \in V; \\ V_j^{(1)} &= \{\vec{v} \in V : v_j = 1\}, \quad V_j^{(-1)} = \{\vec{v} \in V : v_j = -1\}; \\ V_j^{(0)} &= \{\vec{v} : v_i = -1, +1, \forall i \neq j, v_j = 0\}, \quad \forall j = \overline{1, k(\varepsilon)}. \end{aligned}$$

Note that in view of (6.42) and assumption **A4**

$$f(\cdot, \vec{v}) \in \Sigma, \quad \forall \vec{v} \in V. \quad (6.43)$$

Fix some r.n.f.  $\varrho_\varepsilon \in \Omega_\varepsilon$ , satisfying (6.40), and denote by  $\mathcal{B}_\varepsilon$  the random event  $\{\varrho_\varepsilon = a_\varepsilon\}$ . Let us formulate the important auxiliary statement.

There exists  $p_0 > 0$  such that for  $\forall j = \overline{1, k(\varepsilon)}$

$$\frac{1}{2^{k(\varepsilon)-1}} \sum_{\vec{v} \in V_j^{(0)}} \mathbf{P}_{\vec{v}}\{\mathcal{B}_\varepsilon\} \geq p_0 \quad (6.44)$$

for all small enough  $\varepsilon > 0$ .

First, let us show that (6.41) follows from (6.44). Fix some an arbitrary estimator  $\tilde{f}_\varepsilon(\cdot)$ . Then, due to (6.43)

$$\begin{aligned}
R_\varepsilon^{(r)}(\tilde{f}_\varepsilon, \Sigma, \varrho_\varepsilon) &\geq \sup_{f(\cdot) \in \Sigma} \mathbf{E}_f \left( \varrho_\varepsilon^{-1} \|\tilde{f}_\varepsilon(\cdot) - f(\cdot)\|_2 \right)^q I\{\mathcal{B}_\varepsilon\} \\
&\geq \sup_{f(\cdot, \vec{v}), \vec{v} \in V} \mathbf{E}_{f(\cdot, \vec{v})} \left( a_\varepsilon^{-1} \|\tilde{f}_\varepsilon(\cdot) - f(\cdot, \vec{v})\|_2 \right)^q I\{\mathcal{B}_\varepsilon\} \\
&\geq \frac{1}{2^{k(\varepsilon)}} \sum_{\vec{v} \in V} \mathbf{E}_{\vec{v}} \left( a_\varepsilon^{-1} \|\tilde{f}_\varepsilon(\cdot) - f(\cdot, \vec{v})\|_2 \right)^q I\{\mathcal{B}_\varepsilon\} \\
&\geq \left( \frac{1}{2^{k(\varepsilon)}} \sum_{\vec{v} \in V} \mathbf{E}_{\vec{v}} \left( a_\varepsilon^{-2} \|\tilde{f}_\varepsilon(\cdot) - f(\cdot, \vec{v})\|_2^2 \right) I\{\mathcal{B}_\varepsilon\} \right)^{\frac{q}{2}} \triangleq \left( R_\varepsilon(\tilde{f}_\varepsilon) \right)^{\frac{q}{2}}, \quad (6.45)
\end{aligned}$$

where  $\mathbf{E}_{\vec{v}}$  is mathematical expectation w.r.t. measure  $\mathbf{P}_{\vec{v}}$ . Here we used that  $q \geq 2$ ,  $\mathbf{E}|\xi|^s \geq (\mathbf{E}|\xi|)^s, \forall s \geq 1$ , and the fact that

$$\frac{1}{2^{k(\varepsilon)}} \sum_{\vec{v} \in V} = \mathbf{E}_{\vec{v}}, \quad (6.46)$$

where  $\vec{v}$  is the random vector with independent coordinates, taking two values  $+1$  and  $-1$  with probabilities  $\frac{1}{2}$ . Denote for  $\forall j = \overline{1, k(\varepsilon)}$  :  $\Delta_j = [t_j - \delta_\varepsilon, t_j + \delta_\varepsilon]$ ;

$$\begin{aligned}
Z_j^{(1)} &= \frac{d\mathbf{P}_{\vec{v}}}{d\mathbf{P}_{\vec{v}_j(\vec{v})}} (X_\varepsilon(t), 0 \leq t \leq 1), \quad \vec{v} \in V_j^{(1)}; \\
Z_j^{(-1)} &= \frac{d\mathbf{P}_{\vec{v}}}{d\mathbf{P}_{\vec{v}_j(\vec{v})}} (X_\varepsilon(t), 0 \leq t \leq 1), \quad \vec{v} \in V_j^{(-1)},
\end{aligned}$$

where for  $\forall \vec{v} = (v_1, \dots, v_{k(\varepsilon)}) \in V$  the vector  $\vec{v}_j(\vec{v}) \in V_j^{(0)}$  is defined as  $\vec{v}_j(\vec{v}) = (v_1, \dots, v_{j-1}, 0, v_{j+1}, \dots, v_{k(\varepsilon)})$ . Due to Girsanov's formulae

$$Z_j^{(1)} = \exp \left\{ \frac{1}{\varepsilon^2} \int_0^1 (f(t, \vec{v}) - f(t, \vec{v}_j(\vec{v}))) dX_\varepsilon(t) - \frac{1}{2\varepsilon^2} \|f(\cdot, \vec{v})\|_2^2 + \frac{1}{2\varepsilon^2} \|f(\cdot, \vec{v}_j(\vec{v}))\|_2^2 \right\}.$$

Therefore, w.r.t. measure  $\mathbf{P}_{\vec{v}_j(\vec{v})}$ ,  $Z_j^{(1)}$  can be represented as

$$Z_j^{(1)} = \exp \left\{ \frac{\delta_\varepsilon^\beta}{\varepsilon} \int_{\Delta_j} F_j(t) dw(t) - \frac{\delta_\varepsilon^{2\beta}}{2\varepsilon^2} \int_{\Delta_j} F_j^2(t) dt \right\}.$$

Putting,

$$\eta_j = \frac{\int_{\Delta_j} F_j(t) dw(t)}{\sqrt{\int_{\Delta_j} F_j^2(t) dt}}$$

one has

$$Z_j^{(1)} = \exp \left\{ \psi_\varepsilon \eta_j - \frac{1}{2} \psi_\varepsilon^2 \right\}, \quad (6.47)$$

where  $\psi_\varepsilon = y \varepsilon^{\frac{1}{4\beta+1}} \left( \ln \frac{1}{\alpha_\varepsilon} \right)^{\frac{2\beta+1}{8\beta+2}}$ ,  $y = (\sigma \bar{Q})^{\frac{6\beta+1}{8\beta+2}}$ . Note

$$\mathcal{L} \left\{ \eta_j \mid \mathbf{P}_{\vec{v}_j(\vec{v})} \right\} = \mathcal{N}(0, 1), \quad \forall j = \overline{1, k(\varepsilon)}. \quad (6.48)$$

Analogously,

$$Z_j^{(-1)} = \exp \left\{ -\psi_\varepsilon \eta_j - \frac{1}{2} \psi_\varepsilon^2 \right\}. \quad (6.49)$$

From (6.48) we conclude that the distribution of  $Z_j^{(1)}$  and  $Z_j^{(-1)}$  does not depend on  $v_1, \dots, v_{j-1}, v_{j+1}, \dots, v_{k(\varepsilon)}$  and  $j$ . Moreover,  $\psi_\varepsilon \leq y \varepsilon^{\frac{1}{4\beta+1}} \left( a \ln \frac{1}{\varepsilon} \right)^{\frac{2\beta+1}{8\beta+2}} \rightarrow 0$ , as  $\varepsilon \rightarrow 0$ . Hence, for all small enough  $\varepsilon > 0$ , for  $\forall \vec{v} \in V_j^{(0)}$  and  $\forall 0 < \delta < p_0/2$  we have

$$\mathbf{P}_{\vec{v}} \left( \{Z_j^{(1)} < 1 - \delta\} \cup \{Z_j^{(-1)} < 1 - \delta\} \right) \leq \delta. \quad (6.50)$$

Put  $T_\varepsilon(\cdot) = \tilde{f}_\varepsilon(\cdot) - f_0(\cdot)$  and  $D_j = \{Z_j^{(1)} \geq 1 - \delta\} \cap \{Z_j^{(-1)} \geq 1 - \delta\}$ . Continuing (6.45), we get

$$\begin{aligned} R_\varepsilon(\tilde{f}_\varepsilon) &= a_\varepsilon^{-2} \sum_{j=1}^{k(\varepsilon)} \frac{1}{2^{k(\varepsilon)}} \sum_{\vec{v} \in V} \left( \mathbf{E}_{\vec{v}} \left( \int_{\Delta_j} (T_\varepsilon(t) - \bar{f}(t, \vec{v}))^2 dt \right) I\{\mathcal{B}_\varepsilon\} \right) \\ &= a_\varepsilon^{-2} \sum_{j=1}^{k(\varepsilon)} \frac{1}{2^{k(\varepsilon)}} \sum_{\vec{v} \in V_j^{(1)}} \mathbf{E}_{\vec{v}} \left( \int_{\Delta_j} (T_\varepsilon(t) - \bar{f}(t, \vec{v}))^2 dt \right) I\{\mathcal{B}_\varepsilon\} \\ &+ a_\varepsilon^{-2} \sum_{j=1}^{k(\varepsilon)} \frac{1}{2^{k(\varepsilon)}} \sum_{\vec{v} \in V_j^{(-1)}} \mathbf{E}_{\vec{v}} \left( \int_{\Delta_j} (T_\varepsilon(t) - \bar{f}(t, \vec{v}))^2 dt \right) I\{\mathcal{B}_\varepsilon\} \\ &= a_\varepsilon^{-2} \sum_{j=1}^{k(\varepsilon)} \frac{1}{2^{k(\varepsilon)}} \sum_{\vec{v} \in V} \mathbf{E}_{\vec{v}_j(\vec{v})} I\{\mathcal{B}_\varepsilon\} \\ &\quad \left( Z_j^{(1)} \int_{\Delta_j} (T_\varepsilon(t) - F_j(t))^2 dt + Z_j^{(-1)} \int_{\Delta_j} (T_\varepsilon(t) + F_j(t))^2 dt \right) \\ &\geq (1 - \delta) a_\varepsilon^{-2} \sum_{j=1}^{k(\varepsilon)} \frac{1}{2^{k(\varepsilon)}} \sum_{\vec{v} \in V} \mathbf{E}_{\vec{v}_j(\vec{v})} I\{\mathcal{B}_\varepsilon \cap D_j\} \\ &\quad \left( \int_{\Delta_j} \left\{ (T_\varepsilon(t) - F_j(t))^2 + (T_\varepsilon(t) + F_j(t))^2 \right\} dt \right) \\ &\geq (1 - \delta) a_\varepsilon^{-2} \sum_{j=1}^{k(\varepsilon)} \frac{1}{2^{k(\varepsilon)-1}} \sum_{\vec{v} \in V_j^{(0)}} \mathbf{P}_{\vec{v}} \{ \mathcal{B}_\varepsilon \cap D_j \} \int_{\Delta_j} F_j^2(t) dt. \end{aligned} \quad (6.51)$$

Here we used that  $2a^2 + 2b^2 \geq (a - b)^2$ . Since  $\int_{\Delta_j} F_j^2(t) dt = \bar{Q}^2 \sigma^2 \delta_\varepsilon^{2\beta+1}$ , continuing (6.51), we get

$$\begin{aligned} R_\varepsilon(\tilde{f}_\varepsilon) &\geq (1 - \delta) \left( \bar{Q} \sigma \delta_\varepsilon^{\beta+\frac{1}{2}} a_\varepsilon^{-1} \right)^2 \sum_{j=1}^{k(\varepsilon)} \frac{1}{2^{k(\varepsilon)-1}} \sum_{\vec{v} \in V_j^{(0)}} \mathbf{P}_{\vec{v}} \{ \mathcal{B}_\varepsilon \cap D_j \} \\ &\geq (1 - \delta) \left( \bar{Q} \sigma \delta_\varepsilon^{\beta+\frac{1}{2}} a_\varepsilon^{-1} \right)^2 \sum_{j=1}^{k(\varepsilon)} \frac{1}{2^{k(\varepsilon)-1}} \sum_{\vec{v} \in V_j^{(0)}} \left( \mathbf{P}_{\vec{v}} \{ \mathcal{B}_\varepsilon \} - \mathbf{P}_{\vec{v}} \{ D_j^c \} \right). \end{aligned}$$



Due to (6.44) and (6.48)

$$\begin{aligned} \frac{1}{2^{k(\varepsilon)-1}} \sum_{\vec{v} \in V_j^{(0)}} \mathbf{P}_{\vec{v}}\{\mathcal{B}_\varepsilon\} &\geq p_0, \quad \forall j = \overline{1, k(\varepsilon)}; \\ \mathbf{P}_{\vec{v}}\{D_j^c\} &< \delta < \frac{p_0}{2}, \quad \forall j = \overline{1, k(\varepsilon)}, \quad \forall \vec{v} \in V_j^{(0)}, \end{aligned}$$

and we finally arrive

$$\begin{aligned} R_\varepsilon(\tilde{f}_\varepsilon) &\geq (1 - \delta) \left( \bar{Q} \sigma \delta_\varepsilon^{\beta + \frac{1}{2}} a_\varepsilon^{-1} \right)^2 k(\varepsilon) \frac{p_0}{2} \\ &= \text{const} \left( \delta_\varepsilon^{\beta + \frac{1}{4}} a_\varepsilon^{-1} \right)^2 = \text{const} \left( \varphi_\varepsilon(\alpha_\varepsilon) a_\varepsilon^{-1} \right)^2. \end{aligned} \quad (6.52)$$

Noting, that the right side of (6.52) does not depend on the estimator  $\tilde{f}_\varepsilon(\cdot)$ , we have

$$\liminf_{\varepsilon \rightarrow 0} \inf_{\tilde{f}_\varepsilon} R_\varepsilon(\tilde{f}_\varepsilon) \geq \text{const} \liminf_{\varepsilon \rightarrow 0} \left( \varphi_\varepsilon(\alpha_\varepsilon) a_\varepsilon^{-1} \right)^2 = +\infty$$

in view of (6.40). From here and (6.45) we obtain the statement of the theorem. Thus, it remains to prove (6.44).

Fix some  $j = \overline{1, k(\varepsilon)}$  and put

$$Z_\varepsilon = \frac{1}{2^{k(\varepsilon)-1}} \sum_{\vec{v} \in V_j^{(0)}} \frac{d\mathbf{P}_{\vec{v}}}{d\mathbf{P}_0}(X_\varepsilon(t), 0 \leq t \leq 1).$$

Since  $\varrho_\varepsilon \in \Omega_\varepsilon$ , by definition, for  $\forall \delta > 0$

$$\alpha_\varepsilon^{-1} \mathbf{P}_0\{\mathcal{B}_\varepsilon^c\} \leq 1 + \delta,$$

for all small enough  $\varepsilon > 0$ . From here

$$\begin{aligned} \mathcal{P}_\varepsilon &= \frac{1}{2^{k(\varepsilon)-1}} \sum_{\vec{v} \in V_j^{(0)}} \mathbf{P}_{\vec{v}}\{\mathcal{B}_\varepsilon\} = \mathbf{E}_0 Z_\varepsilon I\{\mathcal{B}_\varepsilon\} \\ &\geq \mathbf{E}_0 \left( Z_\varepsilon I\{\mathcal{B}_\varepsilon\} + c \alpha_\varepsilon^{-1} I\{\mathcal{B}_\varepsilon^c\} \right) - c(1 + \delta), \end{aligned} \quad (6.53)$$

where the constant  $c$  will be chosen below. It is clear that

$$\begin{aligned} &\inf_{\mathcal{F}_\varepsilon} \left( Z_\varepsilon I\{\mathcal{F}_\varepsilon\} + c \alpha_\varepsilon^{-1} I\{\mathcal{F}_\varepsilon^c\} \right) \\ &= Z_\varepsilon I\{Z_\varepsilon < c \alpha_\varepsilon^{-1}\} + c \alpha_\varepsilon^{-1} I\{Z_\varepsilon \geq c \alpha_\varepsilon^{-1}\}. \end{aligned}$$

From here, continuing (6.53), one has

$$\begin{aligned} \mathcal{P}_\varepsilon &\geq \mathbf{E}_0 \left( Z_\varepsilon I\{Z_\varepsilon < c \alpha_\varepsilon^{-1}\} + c \alpha_\varepsilon^{-1} I\{Z_\varepsilon \geq c \alpha_\varepsilon^{-1}\} \right) - c(1 + \delta) \\ &\geq \mathbf{E}_0 \left( Z_\varepsilon I\{Z_\varepsilon < c \alpha_\varepsilon^{-1}\} \right) - c(1 + \delta) \\ &= \frac{1}{2^{k(\varepsilon)-1}} \sum_{\vec{v} \in V_j^{(0)}} \mathbf{P}_{\vec{v}}\{Z_\varepsilon < c \alpha_\varepsilon^{-1}\} - c(1 + \delta) \\ &= (1 - c(1 + \delta)) - \frac{1}{2^{k(\varepsilon)-1}} \sum_{\vec{v} \in V_j^{(0)}} \mathbf{P}_{\vec{v}}\{Z_\varepsilon \geq c \alpha_\varepsilon^{-1}\}. \end{aligned}$$

Since,  $Z_\varepsilon \geq 0$ , one can apply Markov inequality. It yields

$$\mathbf{P}_{\vec{\sigma}}\{Z_\varepsilon \geq c\alpha_\varepsilon^{-1}\} \leq \alpha_\varepsilon c^{-1} \mathbf{E}_{\vec{\sigma}} Z_\varepsilon.$$

From here

$$\begin{aligned} \mathcal{P}_\varepsilon &\geq (1 - c(1 + \delta)) - \alpha_\varepsilon c^{-1} \frac{1}{2^{k(\varepsilon)-1}} \sum_{\vec{\sigma} \in V_j^{(0)}} \mathbf{E}_{\vec{\sigma}} Z_\varepsilon \\ &= (1 - c(1 + \delta)) - \alpha_\varepsilon c^{-1} \mathbf{E}_0 Z_\varepsilon^2. \end{aligned} \quad (6.54)$$

Let us now calculate  $\mathbf{E}_0 Z_\varepsilon^2$ . Applying again Girsanov's formulae, we have in view of (6.46)

$$Z_\varepsilon = \prod_{i=1, i \neq j}^{k(\varepsilon)} \left( \frac{1}{2} Z_i^{(1)} + \frac{1}{2} Z_i^{(-1)} \right),$$

where  $Z_i^{(1)}$ ,  $Z_i^{(-1)}$ ,  $i = \overline{1, k(\varepsilon)}$ , are defined by formulas (6.47) and (6.49) respectively. Thus,

$$Z_\varepsilon = \prod_{i=1, i \neq j}^{k(\varepsilon)} \left( \frac{1}{2} \exp\{\psi_\varepsilon \eta_i - \frac{1}{2} \psi_\varepsilon^2\} + \frac{1}{2} \exp\{-\psi_\varepsilon \eta_i - \frac{1}{2} \psi_\varepsilon^2\} \right),$$

and remind that  $\eta_i \sim \mathcal{N}(0, 1)$ ,  $i = \overline{1, k(\varepsilon)}$  w.r.t. measure  $\mathbf{P}_0$ . Note also that  $\mathbf{E}_0 \eta_i \eta_l = 0$ ,  $\forall i, l = \overline{1, k(\varepsilon)}$ ,  $i \neq l$ , because  $\int_0^1 F_i(t) F_l(t) \mathbf{d}t = 0$ , and, therefore,  $\eta_i$ ,  $i = \overline{1, k(\varepsilon)}$ , are independent random variables. Then

$$\begin{aligned} \mathbf{E}_0 Z_\varepsilon^2 &= \left\{ \mathbf{E} \left( \frac{1}{2} \exp\{\psi_\varepsilon \eta_1 - \frac{1}{2} \psi_\varepsilon^2\} + \frac{1}{2} \exp\{-\psi_\varepsilon \eta_1 - \frac{1}{2} \psi_\varepsilon^2\} \right) \right\}^{k(\varepsilon)-1} \\ &= \left( \frac{1}{2} \exp\{\psi_\varepsilon^2\} + \frac{1}{2} \exp\{-\psi_\varepsilon^2\} \right)^{k(\varepsilon)-1} \\ &= \left( 1 + \psi_\varepsilon^4 + \mathcal{O}(\psi_\varepsilon^6) \right)^{k(\varepsilon)-1} = \exp\{k(\varepsilon) \psi_\varepsilon^4\} (1 + o(1)), \end{aligned}$$

because  $\psi_\varepsilon^6 k(\varepsilon) \rightarrow 0$ , as  $\varepsilon \rightarrow 0$ . Since  $k(\varepsilon) \psi_\varepsilon^4 = -\frac{1}{2} \ln \alpha_\varepsilon$ , then  $\mathbf{E}_0 Z_\varepsilon^2 = \alpha_\varepsilon^{-\frac{1}{2}} (1 + o(1))$ . From here and (6.54)

$$\mathcal{P}_\varepsilon \geq 1 - c(1 + \delta) - (\sqrt{\alpha_\varepsilon} c^{-1}) (1 + o(1)).$$

If  $\alpha_\varepsilon \rightarrow 0$ , as  $\varepsilon \rightarrow 0$ , then  $\mathcal{P}_\varepsilon > p_0$  for  $\forall p_0 < 1$ , because  $c$  and  $\delta$  can be chosen arbitrary small numbers. If  $\liminf_{\varepsilon \rightarrow 0} \alpha_\varepsilon > 0$ , then remember that by assumption of the theorem  $\alpha_\varepsilon < 2^{-5}$ ,  $\forall \varepsilon > 0$ . In this case, choosing  $c = \frac{1}{2}$ , we obtain  $\mathcal{P}_\varepsilon \geq \frac{1}{16}$ , if  $\varepsilon$  and  $\delta$  are small enough. This proves (6.44) and completes the proof of the theorem.

Theorem is proved.

6.1.2. *Proof of Lemma 1.* For  $\forall N \geq 1$  and  $\forall k = \overline{1, N}$  denote by  $\mathcal{P}_N^k$  the set of polynomials, defined on interval  $\Delta_k$ , of the form

$$\mathcal{P}_N^k = \left\{ p_N(t), t \in \Delta_k : p_N(t) = \sum_{j=1}^N a_j (N(t - t_{k-1}))^{j-1}; a_j \in \mathbf{R}^1, j = \overline{1, m+1} \right\}.$$

Fix some function  $f(\cdot) \in \Sigma$  and consider the following minimization problem

$$\begin{cases} \int_{\Delta_k} (p_N(t) - f(t))^2 dt \rightarrow \inf \\ p_N(\cdot) \in \mathcal{P}_N^k \end{cases} \quad (6.55)$$

Put  $\bar{\mathbf{a}}^{(k)} = \Lambda^{-1} \mathbf{b}^{(k)}(f)$ . It is easily to see that the polynomial

$$\bar{p}_N^{(k)}(t) = \sum_{j=1}^N \bar{a}_j^{(k)} (N(t - t_{k-1}))^{j-1},$$

is the solution of the problem (6.55), i.e

$$\int_{\Delta_k} (\bar{p}_N^{(k)}(t) - f(t))^2 dt = \inf_{p_N(\cdot) \in \mathcal{P}_N^k} \int_{\Delta_k} (p_N(t) - f(t))^2 dt. \quad (6.56)$$

On the interval  $\Delta_k$  let us define the polynomial  $\tilde{p}_N^{(k)}(\cdot)$  as follows.

$$\tilde{p}_N^{(k)}(t) = \sum_{j=1}^{m+1} \frac{f^{(j-1)}(t_{k-1})}{(j-1)! N^{j-1}} (N(t - t_{k-1}))^{j-1}.$$

Since  $f(\cdot) \in \Sigma$  we have due to Taylor expansion

$$\int_{\Delta_k} (\tilde{p}_N^{(k)}(t) - f(t))^2 dt \leq \left( \frac{Q}{m!} N^{-\beta} \right)^2 \int_{\Delta_k} dt = \frac{1}{N} \left( \frac{Q}{m!} N^{-\beta} \right)^2. \quad (6.57)$$

Since  $\tilde{p}_N^{(k)}(\cdot) \in \mathcal{P}_N^k$  from (6.56) and (6.57) we finally get

$$\|b_N(\cdot, f) - f(\cdot)\|_2^2 = \sum_{k=1}^N \int_{\Delta_k} (\bar{p}_N^{(k)}(t) - f(t))^2 dt \leq \left( \frac{Q}{m!} N^{-\beta} \right)^2.$$

Lemma is proved.

6.1.3. *Proof of Lemma 2.* Put  $T = \frac{C_N}{4+2C_N}$  and  $A = (2 + C_N)(m + 1)$ . To state the lemma it is sufficient to show that  $\forall k = 1, \bar{N}$  and for  $\forall |t| \leq T$

$$\mathbf{E} \exp\{t\eta_1\} \leq \exp\left\{\frac{1}{2}At^2\right\}. \quad (6.58)$$

Then the first two inequalities of the lemma follow from Petrov (1975, ch.2, 4, theorem 17). Moreover,

$$\frac{x^2}{4N(m+1)(1+C_N)} = \frac{x^2}{4N(m+1)} - \frac{x^2 C_N}{4N(m+1)(1+C_N)} = \frac{x^2}{4N(m+1)} - o(1),$$

uniformly w.r.t  $x \in [0, NC_N(m+1)]$ , if  $C_N = o(N^{-\frac{1}{3}})$ , as  $N \rightarrow \infty$  that proves the third statement of the lemma. Thus, one needs to state (6.58). Indeed, for  $t < \frac{1}{2}$

$$\mathbf{E} \exp\{t\eta_1\} = \left( \exp\{t\} \sqrt{1-2t} \right)^{-(m+1)}.$$

Simplest direct calculations show that for  $\forall |t| \leq T$

$$(1-2t)^{-\frac{(m+1)}{2}} \exp\left\{-(m+1)t - \frac{1}{2}At^2\right\} \leq 1,$$

that proves (6.58).

Lemma is proved.

6.1.4. *Proof of Lemma 4.* Due to assumption **A1**, the set  $\Theta$  is bounded and, therefore, there exists  $P > 0$  such that  $\Theta \in [-P, P]^s$ . Fix some  $b > 0$  (the final choice of  $b$  will be done below). Put  $p_i = i\varepsilon^b$ ,  $i = 0, \pm 1, \dots, \pm I(\varepsilon)$ , where  $I(\varepsilon) = [P\varepsilon^{-b}]$ , and let  $S_\varepsilon$  be the  $\varepsilon^b$ -net in the interval  $[-P, P]$ , i.e.  $S_\varepsilon = \{p_i, i = \pm 1, \dots, \pm I(\varepsilon)\}$ . Let  $S_\varepsilon^{(s)} = S_\varepsilon \times \dots \times S_\varepsilon$  and denote  $\Theta_\varepsilon = \Theta \cap S_\varepsilon^{(s)}$ .  $\Theta_\varepsilon$  is the net in the set  $\Theta$  and, obviously,  $|\Theta_\varepsilon| \leq (2P\varepsilon^{-b})^s$ . For  $\forall x \in [-P, P]$  define the transformation  $u(x) : x \rightarrow S_\varepsilon$  as follows:  $u(x) < x \leq u(x) + \varepsilon^b$ . For  $\forall \vartheta \in \Theta$  put  $u^{(s)}(\vartheta) = (u(\vartheta_1), \dots, u(\vartheta_s))$ . Clearly that  $u^{(s)}(\vartheta) \in \Theta_\varepsilon$  for  $\forall \vartheta \in \Theta$ ,  $u^{(s)}(\vartheta)$  is uniquely defined, and

$$\|u^{(s)}(\vartheta) - \vartheta\| \leq \varepsilon^b \sqrt{s}. \quad (6.59)$$

Put, at last,  $\bar{\vartheta}_\varepsilon = u^{(s)}(\hat{\vartheta}_\varepsilon)$  and note that  $\bar{\vartheta}_\varepsilon$  is measurable w.r.t. the observation  $(X_\varepsilon(t), 0 \leq t \leq 1)$  in view of the definition of  $u^{(s)}(\cdot)$ . Denote for  $\forall \vartheta \in \Theta$

$$\begin{aligned} \Delta_\varepsilon^2(f, \vartheta) &= \frac{1}{N_\varepsilon} \sum_{k=1}^{N_\varepsilon} \|\mathbf{c}^{(k)}(f) - \mathbf{y}^{(k)}(\vartheta)\|^2; \\ z_\varepsilon(f, \vartheta) &= \frac{2\varepsilon}{\sqrt{N_\varepsilon}} \sum_{k=1}^{N_\varepsilon} (\mathbf{c}^{(k)}(f) - \mathbf{y}^{(k)}(\vartheta))' \eta^{(k)}(\varepsilon). \end{aligned}$$

Later on we will use the standard trick. We show that the statement of the lemma holds, if to replace  $\hat{\vartheta}_\varepsilon$  by  $\bar{\vartheta}_\varepsilon$ . In other words, we prove required inequality on the net  $\Theta_\varepsilon$  and then show that the differences between  $\Delta_\varepsilon^2(f) = \Delta_\varepsilon^2(f, \hat{\vartheta}_\varepsilon)$ ,  $z_\varepsilon(f) = z_\varepsilon(f, \hat{\vartheta}_\varepsilon)$ , and  $\Delta_\varepsilon^2(f, \bar{\vartheta})$ ,  $z_\varepsilon(f, \bar{\vartheta})$  are negligible. First, let us suppose that

$$\begin{aligned} \sup_{f(\cdot) \in \Sigma} \mathbf{P}_f \left\{ |z_\varepsilon(f, \bar{\vartheta})| \geq \frac{\delta_1}{2} \Delta_\varepsilon^2(f, \bar{\vartheta}) ; \Delta_\varepsilon^2(f, \bar{\vartheta}) \geq \left( \frac{\delta_2}{2} \varphi_\varepsilon(\alpha_\varepsilon) \right)^2 \right\} \\ \leq v_3 \exp \left\{ -v_4 \varepsilon^{-\mu} \right\} \end{aligned} \quad (6.60)$$

and conclude from here the statement of the lemma. We will use the notation  $L_1, L_2, \dots$  for the absolute constants. Indeed, since we consider the functions  $f(\cdot) \in \Sigma_1(L)$ , then

$$\|\mathbf{c}^{(k)}(f)\| \leq L_1, \quad \|\mathbf{y}^{(k)}(\vartheta)\| \leq L_1. \quad (6.61)$$

From (6.61)

$$\begin{aligned} |\Delta_\varepsilon^2(f) - \Delta_\varepsilon^2(f, \bar{\vartheta}_\varepsilon)| &\leq L_2 \frac{1}{N_\varepsilon} \sum_{k=1}^{N_\varepsilon} \|\mathbf{y}^{(k)}(\hat{\vartheta}_\varepsilon) - \mathbf{y}^{(k)}(\bar{\vartheta})\|^2 \\ &\leq L_3 \|\hat{\vartheta}_\varepsilon - \bar{\vartheta}_\varepsilon\|^\nu \leq L_3 s^{\frac{\nu}{2}} \varepsilon^{b\nu}. \end{aligned} \quad (6.62)$$

Choosing  $b\nu = 2$ , we get from (6.62) for all small enough  $\varepsilon > 0$  the following inclusion

$$\left\{ \Delta_\varepsilon^2(f) \geq (\delta_2 \varphi_\varepsilon(\alpha_\varepsilon))^2 \right\} \subseteq \left\{ \Delta_\varepsilon^2(f, \bar{\vartheta}_\varepsilon) \geq \left( \frac{\delta_2}{2} \varphi_\varepsilon(\alpha_\varepsilon) \right)^2 \right\}. \quad (6.63)$$

Quite similarly

$$|z_\varepsilon(f) - z_\varepsilon(f, \bar{\vartheta}_\varepsilon)| \leq L_4 \sqrt{N_\varepsilon} \varepsilon^{b\nu+1} \left| \frac{1}{N_\varepsilon} \sum_{k=1}^{N_\varepsilon} |\tilde{\eta}_k(\varepsilon)| \right|, \quad (6.64)$$

where  $\tilde{\eta}_k(\varepsilon) = \sum_{j=1}^{m+1} |\eta_j^{(k)}(\varepsilon)|$ . Taking into account, that  $\eta_j^{(k)}(\varepsilon)$ ,  $k = \overline{1, N_\varepsilon}$ ,  $j = \overline{1, m+1}$ , are i.i. normally distributed random variables, from (6.64) with  $b\nu = 2$  one has

$$\sup_{f(\cdot) \in \Sigma} \mathbf{P}_f \left\{ |z_\varepsilon(f) - z_\varepsilon(f, \bar{\vartheta}_\varepsilon)| \geq \varepsilon^2 \right\} \leq L_5 \exp\{-L_6 \varepsilon^{-2}\}.$$

The last expression together with (6.60) and (6.63) leads to the statement of the lemma. Thus, it remains to prove (6.60).

Denote  $\Theta_\varepsilon^+(f) = \{\vartheta \in \Theta_\varepsilon : \Delta_\varepsilon(f, \vartheta) \geq \frac{\delta}{2} \varphi_\varepsilon(\alpha_\varepsilon)\}$  and set for  $\forall f(\cdot) \in \Sigma$  and for  $\forall \vartheta \in \Theta_\varepsilon^+(f)$

$$\tilde{z}_\varepsilon(f, \vartheta) = \frac{z_\varepsilon(f, \vartheta)}{2\varepsilon \Delta_\varepsilon(f, \vartheta)}.$$

Note that  $\tilde{z}_\varepsilon(f, \vartheta) \sim \mathcal{N}(0, 1)$  for  $\forall \vartheta \in \Theta$  and for  $\forall f(\cdot) \in \Sigma$ . From here we get  $\forall f(\cdot) \in \Sigma$

$$\begin{aligned} & \mathbf{P}_f \left\{ |z_\varepsilon(f, \bar{\vartheta}_\varepsilon)| \geq \frac{\delta_1}{2} \Delta_\varepsilon^2(f, \bar{\vartheta}_\varepsilon) ; \Delta_\varepsilon^2(f, \bar{\vartheta}_\varepsilon) > \left( \frac{\delta_2}{2} \varphi_\varepsilon(\alpha_\varepsilon) \right)^2 \right\} \\ &= \sum_{\vartheta \in \Theta_\varepsilon^+(f)} \mathbf{P}_f \left\{ |z_\varepsilon(f, \vartheta)| \geq \frac{\delta_1}{2} \Delta_\varepsilon^2(f, \vartheta) ; \bar{\vartheta}_\varepsilon = \vartheta \right\} \\ &\leq \sum_{\vartheta \in \Theta_\varepsilon^+(f)} \mathbf{P}_f \left\{ |\tilde{z}_\varepsilon(f, \vartheta)| \geq \frac{\delta_1 \delta_2}{8\varepsilon} \varphi_\varepsilon(\alpha_\varepsilon) \right\} \\ &\leq |\Theta_\varepsilon^+(f)| \exp \left\{ -L_7 \left( \frac{\varphi_\varepsilon(\alpha_\varepsilon)}{\varepsilon} \right)^2 \right\} \leq |\Theta_\varepsilon| \exp \left\{ -L_7 \left( \frac{\varphi_\varepsilon(\alpha_\varepsilon)}{\varepsilon} \right)^2 \right\} \\ &\leq L_8 \varepsilon^{-\frac{2s}{\nu}} \exp \left\{ -L_7 \left( \ln \frac{1}{\alpha_\varepsilon} \right)^{\frac{\beta}{4\beta+1}} \varepsilon^{-\frac{2}{4\beta+1}} \right\}. \end{aligned}$$

This proves (6.60) with any  $\mu < \frac{2}{4\beta+1}$  and, therefore, completes the proof of the lemma.

Lemma is proved.

**6.2. Proof of Theorem 2.** Put for  $\forall k = \overline{1, N_\varepsilon}$ ,  $k = \overline{1, \tilde{N}_\varepsilon}$  and for  $\forall f(\cdot) \in \Sigma$

$$\begin{aligned}
b_j^{(k)}(f) &= N_\varepsilon^j \int_{\Delta_k} (t - t_{k-1})^{j-1} f(t) \mathbf{d}t; \quad j = \overline{1, l+1}; \\
(\mathbf{b}^{(k)}(f))' &= (b_1^{(k)}(f), \dots, b_{l+1}^{(k)}(f)); \quad \mathbf{c}^{(k)}(f) = \Lambda_l^{-\frac{1}{2}} \mathbf{b}^{(k)}(f); \\
b_\varepsilon(t, f) &= \sum_{k=1}^{N_\varepsilon} (\mathbf{c}^{(k)}(f))' \Lambda_l^{-\frac{1}{2}} \mathbf{d}^{(k)}(t) I\{t \in \Delta_k\}; \\
\tilde{b}_j^{(k)}(f) &= \tilde{N}_\varepsilon^j \int_{\tilde{\Delta}_k} (t - \tilde{t}_{k-1})^{j-1} f(t) \mathbf{d}t; \quad j = \overline{1, l+1}; \\
(\tilde{\mathbf{b}}^{(k)}(f))' &= (\tilde{b}_1^{(k)}(f), \dots, \tilde{b}_{l+1}^{(k)}(f)); \quad \tilde{\mathbf{c}}^{(k)}(f) = \Lambda_l^{-\frac{1}{2}} \tilde{\mathbf{b}}^{(k)}(f); \\
\tilde{b}_\varepsilon(t, f) &= \sum_{k=1}^{\tilde{N}_\varepsilon} (\tilde{\mathbf{c}}^{(k)}(f))' \Lambda_l^{-\frac{1}{2}} \tilde{\mathbf{d}}^{(k)}(t) I\{t \in \tilde{\Delta}_k\} \\
\xi_j^{(k)}(\varepsilon) &= N_\varepsilon^{j-\frac{1}{2}} \int_{\Delta_k} (t - t_{k-1})^{j-1} \mathbf{d}w(t); \quad j = \overline{1, l+1}; \\
(\boldsymbol{\xi}^{(k)}(\varepsilon))' &= (\xi_1^{(k)}(\varepsilon), \dots, \xi_{l+1}^{(k)}(\varepsilon)); \quad \boldsymbol{\eta}^{(k)}(\varepsilon) = \Lambda_l^{-\frac{1}{2}} \boldsymbol{\xi}^{(k)}(\varepsilon); \\
\eta_\varepsilon(t) &= \sum_{k=1}^{N_\varepsilon} (\boldsymbol{\eta}^{(k)}(\varepsilon))' \Lambda_l^{-\frac{1}{2}} \mathbf{d}^{(k)}(t) I\{t \in \Delta_k\}; \\
\tilde{\xi}_j^{(k)}(\varepsilon) &= \tilde{N}_\varepsilon^{j-\frac{1}{2}} \int_{\tilde{\Delta}_k} (t - \tilde{t}_{k-1})^{j-1} \mathbf{d}w(t); \quad j = \overline{1, l+1}; \\
(\tilde{\boldsymbol{\xi}}^{(k)}(\varepsilon))' &= (\tilde{\xi}_1^{(k)}(\varepsilon), \dots, \tilde{\xi}_{l+1}^{(k)}(\varepsilon)); \quad \tilde{\boldsymbol{\eta}}^{(k)}(\varepsilon) = \Lambda_l^{-\frac{1}{2}} \tilde{\boldsymbol{\xi}}^{(k)}(\varepsilon); \\
\tilde{\eta}_\varepsilon(t) &= \sum_{k=1}^{\tilde{N}_\varepsilon} (\tilde{\boldsymbol{\eta}}^{(k)}(\varepsilon))' \Lambda_l^{-\frac{1}{2}} \tilde{\mathbf{d}}^{(k)}(t) I\{t \in \tilde{\Delta}_k\}.
\end{aligned}$$

Note that by definition  $\hat{f}_\varepsilon(\cdot) = b_\varepsilon(\cdot, f) + \varepsilon \sqrt{N_\varepsilon} \eta_\varepsilon(\cdot)$ ,  $\tilde{f}_\varepsilon(\cdot) = \tilde{b}_\varepsilon(\cdot, f) + \varepsilon \sqrt{\tilde{N}_\varepsilon} \tilde{\eta}_\varepsilon(\cdot)$  and this is the standard decomposition of linear estimators onto "smoother" and "stochastic term".

**Lemma 5.** For  $\forall f(\cdot) \in \Sigma$  and for  $\forall \varepsilon \in (0, 1)$  such that  $\tilde{N}_\varepsilon \geq 1$ ,  $N_\varepsilon \geq 2$  the following decomposition holds.

$$\hat{d}_\varepsilon = \|b_\varepsilon(\cdot, f) - \tilde{b}_\varepsilon(\cdot, f)\|_2^2 + 2\varepsilon \sqrt{N_\varepsilon} \langle (b_\varepsilon(\cdot, f) - \tilde{b}_\varepsilon(\cdot, f)), \eta_\varepsilon(\cdot) \rangle + \varepsilon^2 (S_\varepsilon - \tilde{S}_\varepsilon),$$

where

$$\begin{aligned}
\langle (b_\varepsilon(\cdot, f) - \tilde{b}_\varepsilon(\cdot, f)), \eta_\varepsilon(\cdot) \rangle &= \int_0^1 (b_\varepsilon(t, f) - \tilde{b}_\varepsilon(t, f)) \eta_\varepsilon(t) \mathbf{d}t; \\
S_\varepsilon &= \sum_{k=1}^{N_\varepsilon} \sum_{j=1}^{l+1} \left( (\eta_j^{(k)}(\varepsilon))^2 - 1 \right); \quad \tilde{S}_\varepsilon = \sum_{k=1}^{\tilde{N}_\varepsilon} \sum_{j=1}^{l+1} \left( (\tilde{\eta}_j^{(k)}(\varepsilon))^2 - 1 \right);
\end{aligned}$$

6.2.1. *Proof of Theorem 2. I(Upper bound). 1.* Set  $\varphi_\varepsilon = (\lambda\psi_\varepsilon(\alpha_\varepsilon))^2 + (\lambda_1\varphi_\varepsilon(\gamma) + \lambda_2\tau_\varepsilon)^2$ . Let us first prove that  $\hat{\rho}_\varepsilon \in \Omega_\varepsilon$ . By definition, it is required to show that

$$\begin{aligned} \limsup_{\varepsilon \rightarrow 0} \alpha_\varepsilon^{-1} \sup_{f(\cdot) \in \Sigma_0} \mathbf{P}_f \{ \hat{\rho}_\varepsilon = \varphi_\varepsilon(\beta) \} = \\ \limsup_{\varepsilon \rightarrow 0} \alpha_\varepsilon^{-1} \sup_{f(\cdot) \in \Sigma_0} \mathbf{P}_f \{ \hat{d}_\varepsilon > \varphi_\varepsilon \}. \end{aligned} \quad (6.65)$$

Fix some function  $f(\cdot) \in \Sigma_0$ . Since  $f(\cdot) \in H(\gamma, P)$ , the statement of Lemma 1 remains valid, if to replace  $\beta$ ,  $m$ ,  $Q$  by  $\gamma$ ,  $l$ ,  $P$  respectively. Then  $\forall f(\cdot) \in \Sigma_0$  we have by triangle inequality and Lemma 1

$$\begin{aligned} \|b_\varepsilon(\cdot, f) - \tilde{b}_\varepsilon(\cdot, f)\|_2 &\leq \|b_\varepsilon(\cdot, f) - f(\cdot)\|_2 + \|\tilde{b}_\varepsilon(\cdot, f) - f(\cdot)\|_2 \\ &\leq \frac{P}{l!} (N_\varepsilon^{-\gamma} + \tilde{N}_\varepsilon^{-\gamma}) = \lambda_1\varphi_\varepsilon(\gamma) + \lambda_2\tau_\varepsilon. \end{aligned} \quad (6.66)$$

Set  $\forall f(\cdot) \in \Sigma$

$$\zeta_\varepsilon(f) = 2\varepsilon\sqrt{N_\varepsilon} \langle (b_\varepsilon(\cdot, f) - \tilde{b}_\varepsilon(\cdot, f)), \eta_\varepsilon(\cdot) \rangle. \quad (6.67)$$

It is easily to check that for each function  $f(\cdot)$  and  $\forall \varepsilon > 0$

$$\zeta_\varepsilon(f) \sim \mathcal{N} \left( 0, 4\varepsilon^2 \|b_\varepsilon(\cdot, f) - \tilde{b}_\varepsilon(\cdot, f)\|_2^2 \right). \quad (6.68)$$

From (6.66)  $\forall f(\cdot) \in \Sigma_0$

$$\mathbf{E}\zeta_\varepsilon^2(f) \leq 4\varepsilon^2 (\lambda_1\varphi_\varepsilon(\gamma) + \lambda_2\tau_\varepsilon)^2$$

and we see that acrossing term  $\zeta_\varepsilon(f)$  is negligibly small in comparison with threshold  $\varphi_\varepsilon$ , as  $\varepsilon \rightarrow 0$ . Note also that sum  $\varepsilon^2 \tilde{S}_\varepsilon$  satisfies the assumptions of Lemma 2, and we can state the following inequality.

$$\sup_{f(\cdot) \in \Sigma} \mathbf{P}_f \left\{ \varepsilon^2 \tilde{S}_\varepsilon > \varepsilon^{\mu_1} \right\} \leq \exp \left\{ -\varepsilon^{-\mu_2} \right\} (1 + o(1)), \quad (6.69)$$

as  $\varepsilon \rightarrow 0$ , where the constants  $\mu_1 > 0$  and  $\mu_2 > 0$  depend only on  $\gamma$  and  $\beta$ . Here we used the fact that the distribution of the random variable  $\varepsilon^2 \tilde{S}_\varepsilon$  does not depend on  $f(\cdot)$ . Thus, we get from (6.66), (6.68), (6.69) and Lemma 5

$$\mathbf{P}_f \{ \hat{d}_\varepsilon \geq \varphi_\varepsilon \} = \mathbf{P}_f \left\{ S_\varepsilon \geq \left( \frac{\lambda\psi(\alpha_\varepsilon)}{\varepsilon} \right)^2 \right\} (1 + o(1))$$

uniformly w.r.t.  $f(\cdot) \in \Sigma_0$ , as  $\varepsilon \rightarrow 0$ . Noting again, that distribution of the random variable  $S_\varepsilon$  does not depend on  $f(\cdot)$ , and applying Lemma 2 with  $x = \left( \frac{\lambda\psi(\alpha_\varepsilon)}{\varepsilon} \right)^2$  to the sum  $S_\varepsilon$ , we arrive to the required inequality (6.65), because

$$\left( \frac{\lambda\psi_\varepsilon(\alpha_\varepsilon)}{\varepsilon} \right)^4 \frac{1}{4N_\varepsilon(l+1)} = \ln \frac{1}{\alpha_\varepsilon}.$$

**2.** Now let us prove the inequality (5.12). We will apply the similar approach, which we used to prove (5.10) in Theorem 1.

Set  $\forall f(\cdot) \in \Sigma$

$$R_\varepsilon^{(1)}(f) = \mathbf{E}_f \left( \varphi_\varepsilon^{-1}(\alpha_\varepsilon) \|\tilde{f}_\varepsilon(\cdot) - f(\cdot)\|_2 \right)^q I\{\mathcal{A}_\varepsilon\}; \quad (6.70)$$

$$R_\varepsilon^{(2)}(f) = \mathbf{E}_f \left( \varphi_\varepsilon^{-1}(\beta) \|\hat{f}_\varepsilon(\cdot, \beta, Q) - f(\cdot)\|_2 \right)^q I\{\mathcal{A}_\varepsilon^c\}. \quad (6.71)$$

As it follows from (6.19) in order to obtain (5.12) it is enough to find uniform upper estimates for  $R_\varepsilon^{(1)}(\cdot)$  and  $R_\varepsilon^{(2)}(\cdot)$ . Since

$$\sup_{f(\cdot) \in \Sigma} R_\varepsilon^{(2)}(f) \leq \sup_{f_\varepsilon^* \in \Sigma} \mathbf{E}_f \left( \varphi_\varepsilon^{-1}(\beta) \|\hat{f}_\varepsilon(\cdot, \beta, Q) - f(\cdot)\|_2 \right)^q,$$

set  $\Sigma$  and the estimator  $\hat{f}_\varepsilon(\cdot, \beta, Q)$  are the same as in Theorem 1, we have from (6.22)

$$\limsup_{\varepsilon \rightarrow 0} \sup_{f(\cdot) \in \Sigma} R_\varepsilon^{(2)}(f) \leq \left( \frac{2\beta + 1}{2\beta^{\frac{2\beta}{2\beta+1}}} \left( \frac{(m+1)^\beta Q}{m!} \right)^{\frac{1}{2\beta+1}} \right)^q. \quad (6.72)$$

Let us now start with the estimation of  $R_\varepsilon^{(1)}(\cdot)$ . First, note that

$$\begin{aligned} \|\tilde{f}_\varepsilon(\cdot) - f(\cdot)\|_2 &= \|\tilde{b}_\varepsilon(\cdot, f) + \tilde{\eta}_\varepsilon(\cdot) - f(\cdot)\|_2 \\ &\leq \|\tilde{b}_\varepsilon(\cdot, f) - b_\varepsilon(\cdot, f)\|_2 + \|b_\varepsilon(\cdot, f) - f(\cdot)\|_2 + \|\tilde{\eta}_\varepsilon(\cdot)\|_2 \\ &= B_\varepsilon(f) + \|b_\varepsilon(\cdot, f) - f(\cdot)\|_2 + \varepsilon \left( \tilde{S}_\varepsilon + \tilde{N}_\varepsilon(l+1) \right)^{\frac{1}{2}} \\ &\leq B_\varepsilon(f) + \frac{Q}{m!} N_\varepsilon^{-\beta} + \varepsilon \left( \tilde{S}_\varepsilon + \tilde{N}_\varepsilon(l+1) \right)^{\frac{1}{2}}, \end{aligned} \quad (6.73)$$

where  $B_\varepsilon(f) = \|\tilde{b}_\varepsilon(\cdot, f) - b_\varepsilon(\cdot, f)\|_2$ . Here we used Lemma 1. Denote  $\bar{d}_\varepsilon = B_\varepsilon^2(f) + \zeta_\varepsilon(f) + \varepsilon^2 S_\varepsilon$ . As it follows from Lemma 5  $\hat{d}_\varepsilon = \bar{d}_\varepsilon - \varepsilon^2 \tilde{S}_\varepsilon$ . Putting  $\forall f(\cdot) \in \Sigma$

$$\bar{R}_\varepsilon^{(1)}(f) = \left( \varphi_\varepsilon^{-1}(\alpha_\varepsilon) \left( B_\varepsilon(f) + \frac{Q}{m!} N_\varepsilon^{-\beta} + \varepsilon \sqrt{\tilde{N}_\varepsilon(l+1)} \right) \right)^q \mathbf{P}_f\{\bar{d}_\varepsilon < \varphi_\varepsilon\},$$

we can state in view (6.69) and (6.73) that

$$\limsup_{\varepsilon \rightarrow 0} \sup_{f(\cdot) \in \Sigma} R_\varepsilon^{(1)}(f) \leq \limsup_{\varepsilon \rightarrow 0} \sup_{f(\cdot) \in \Sigma} \bar{R}_\varepsilon^{(1)}(f).$$

Thus, it is enough to estimate  $\bar{R}_\varepsilon^{(1)}(\cdot)$ . Fix some small  $\delta > 0$  (later on  $\delta \rightarrow 0$ ) and denote

$$\Sigma_\varepsilon^{(1)} = \left\{ f(\cdot) \in \Sigma : B_\varepsilon(f) \leq \sqrt{\varphi_\varepsilon \left( \frac{1+\delta}{1-\delta} \right)} \right\}.$$

Then

$$\begin{aligned} \bar{R}_\varepsilon^{(1,1)} &\triangleq \sup_{f(\cdot) \in \Sigma_\varepsilon^{(1)}} \bar{R}_\varepsilon^{(1)}(f) \\ &\leq \left( \varphi_\varepsilon^{-1}(\alpha_\varepsilon) \left( \sqrt{\varphi_\varepsilon \left( \frac{1+\delta}{1-\delta} \right)} + \frac{Q}{m!} N_\varepsilon^{-\beta} + \varepsilon \sqrt{\tilde{N}_\varepsilon(l+1)} \right) \right)^q. \end{aligned} \quad (6.74)$$



Set  $\Sigma_\varepsilon^{(2)} = \Sigma \setminus \Sigma_\varepsilon^{(1)}$ . Due to (6.68) there exist some constant  $\mu_3 > 0$  such that

$$\sup_{f(\cdot) \in \Sigma_\varepsilon^{(2)}} \mathbf{P} \{ |\zeta_\varepsilon(f)| \geq \delta B_\varepsilon^2(f) \} \leq \exp \{ -\varepsilon^{-\mu_3} \} (1 + o(1)).$$

Hence, for all small enough  $\varepsilon > 0$  and  $\forall f(\cdot) \in \Sigma_\varepsilon^{(2)}$

$$\begin{aligned} \bar{R}_\varepsilon^{(1)}(f) &\leq \left( \varphi_\varepsilon^{-1}(\alpha_\varepsilon) \left( B_\varepsilon(f) + \frac{Q}{m!} N_\varepsilon^{-\beta} + \varepsilon \sqrt{\tilde{N}_\varepsilon(l+1)} \right) \right)^q \\ &\times \mathbf{P}_f \{ B_\varepsilon^2(f)(1-\delta) + \varepsilon^2 S_\varepsilon < \varphi_\varepsilon \}. \end{aligned} \quad (6.75)$$

Now let us consider 3 cases.

**a.**  $\beta < \gamma < 2\beta$ . In this case  $\varphi_\varepsilon(\alpha_\varepsilon) = \varphi_\varepsilon(\gamma)$  and  $\varphi_\varepsilon = (\lambda_1 \varphi_\varepsilon(\gamma))^2 (1 + o(1))$ . We have from (6.74)

$$\begin{aligned} \limsup_{\varepsilon \rightarrow 0} \bar{R}_\varepsilon^{(1,1)} &\leq \lim_{\delta \rightarrow 0} \left( \sqrt{\frac{1+\delta}{1-\delta}} \lambda_1 + \left( \frac{2P\gamma}{\sqrt{l+1}!} \right)^{\frac{1}{2\gamma+1}} \sqrt{l+1} \right)^q \\ &= \left( \frac{2\gamma+1}{(2\gamma)^{\frac{2\gamma}{2\gamma+1}}} \left( \frac{(l+1)^\gamma P}{l!} \right)^{\frac{1}{2\gamma+1}} \right)^q. \end{aligned} \quad (6.76)$$

For all  $f(\cdot) \in \Sigma_\varepsilon^{(2)}$

$$\mathbf{P}_f \{ B_\varepsilon^2(f)(1-\delta) + \varepsilon^2 S_\varepsilon \leq \varphi_\varepsilon \} \leq \mathbf{P}_f \{ \varepsilon^2 S_\varepsilon \leq -\delta \varphi_\varepsilon \}.$$

In view of Lemma 2

$$\sup_{f(\cdot) \in \Sigma_\varepsilon^{(2)}} \mathbf{P}_f \{ \varepsilon^2 S_\varepsilon \leq -\delta \varphi_\varepsilon \}$$

is exponentially small for all  $\delta > 0$ , as  $\varepsilon \rightarrow 0$ , and, therefore,

$$\limsup_{\varepsilon \rightarrow 0} \sup_{f(\cdot) \in \Sigma_\varepsilon^{(2)}} \bar{R}_\varepsilon^{(1)}(f) = 0. \quad (6.77)$$

Here we also used that  $|B_\varepsilon(f)| < L_1, \forall f(\cdot) \in \Sigma$ , for some  $0 < L_1 < \infty$ . It remains to note that

$$\sup_{f(\cdot) \in \Sigma} \bar{R}_\varepsilon^{(1)}(f) = \max \left\{ \sup_{f(\cdot) \in \Sigma_\varepsilon^{(1)}} \bar{R}_\varepsilon^{(1)}(f); \sup_{f(\cdot) \in \Sigma_\varepsilon^{(2)}} \bar{R}_\varepsilon^{(1)}(f) \right\}$$

and we have from (6.76) and (6.77)

$$\limsup_{\varepsilon \rightarrow 0} \sup_{f(\cdot) \in \Sigma} \bar{R}_\varepsilon^{(1)}(f) \leq \left( \frac{2\gamma+1}{(2\gamma)^{\frac{2\gamma}{2\gamma+1}}} \left( \frac{(l+1)^\gamma P}{l!} \right)^{\frac{1}{2\gamma+1}} \right)^q.$$

Last expression together with (6.72) completes the proof of (5.12) for the case  $\beta < \gamma < 2\beta$ .

b.  $\gamma > 2\beta$ . In this case  $\varphi_\varepsilon(\alpha_\varepsilon) = \psi_\varepsilon(\alpha_\varepsilon)$  and  $\varphi_\varepsilon = (\lambda\psi_\varepsilon(\alpha_\varepsilon))^2(1 + o(1))$ . Set  $c_\varepsilon = -q \ln \varepsilon$  and denote for every  $c \in [-\delta \ln \alpha_\varepsilon, c_\varepsilon]$

$$\begin{aligned}\Sigma_\varepsilon(c) &= \left\{ f(\cdot) \in \Sigma : B_\varepsilon^2(f) = \left( \frac{1 + \sqrt{\frac{c}{\ln \frac{1}{\alpha_\varepsilon}}}}{1 - \delta} \right) \varphi_\varepsilon \right\}; \\ \Sigma_\varepsilon^{(3)} &= \left\{ f(\cdot) \in \Sigma : B_\varepsilon^2(f) > \left( \frac{1 + \sqrt{\frac{c_\varepsilon}{\ln \frac{1}{\alpha_\varepsilon}}}}{1 - \delta} \right) \varphi_\varepsilon \right\}.\end{aligned}$$

Then, due to Lemma 2, quite similarly (6.77) and (6.32)

$$\limsup_{\varepsilon \rightarrow 0} \sup_{f(\cdot) \in \Sigma_\varepsilon^{(3)}} \bar{R}_\varepsilon^{(1)}(f) = 0. \quad (6.78)$$

We also obtain from Lemma 2

$$\bar{R}_\varepsilon^{(1)}(f) \leq \left( \left( \left( \frac{1 + \sqrt{\frac{c}{\ln \frac{1}{\alpha_\varepsilon}}}}{1 - \delta} \right)^{\frac{1}{2}} \lambda + \frac{Q}{m!} Z_1^{4\beta} \right)^q \exp\{-c\} \right) (1 + o(1))$$

uniformly w.r.t.  $f(\cdot) \in \Sigma_\varepsilon(c)$  and  $c \in [-\delta \ln \alpha_\varepsilon, c_\varepsilon]$ . Therefore,

$$\begin{aligned}& \sup_{c \in [-\delta \ln \alpha_\varepsilon, c_\varepsilon]} \sup_{f(\cdot) \in \Sigma_\varepsilon(c)} \bar{R}_\varepsilon^{(1)}(f) \\ & \leq \left( \sup_{c \in (0, \infty)} \left( \left( \frac{1 + \sqrt{\frac{c}{\ln \frac{1}{\alpha_\varepsilon}}}}{1 - \delta} \right)^{\frac{1}{2}} \lambda + \frac{Q}{m!} Z_1^{4\beta} \right)^q \exp\{-c\} \right) (1 + o(1)).\end{aligned} \quad (6.79)$$

We also have from (6.74)

$$\sup_{f(\cdot) \in \Sigma_\varepsilon^{(1)}} \bar{R}_\varepsilon^{(1)}(f) \leq \left( \sqrt{\frac{1 + \delta}{1 - \delta}} \lambda + \frac{Q}{m!} Z_1^{4\beta} \right)^q (1 + o(1)). \quad (6.80)$$

Tending, first,  $\varepsilon$  to zero, and, then  $\delta$  to zero, we obtain from (6.78), (6.79) and (6.80)

$$\begin{aligned}& \limsup_{\varepsilon \rightarrow 0} \sup_{f(\cdot) \in \Sigma} \bar{R}_\varepsilon^{(1)}(f) \\ & \leq \sup_{c \in (0, \infty)} \left( \frac{(4(l+1))^{\frac{1}{4}}}{Z_1} \left( 1 + \sqrt{\frac{c}{\ln \frac{1}{\alpha}}} \right)^{\frac{1}{2}} + \frac{Q}{m!} Z_1^{4\beta} \right)^q \exp\{-c\} = G_\alpha(Z_1).\end{aligned}$$

This, together with (6.72), completes the proof of (5.12) for the case  $\gamma > 2\beta$ .

c.  $\gamma = 2\beta$ . We omit the proof of (5.12) for  $\gamma = 2\beta$ , because it is just the repetition of the proof for the previous case. Moreover, the constant, which bounds the risk, is rather cumbersome.

**II(Lower bound).** We do not need to prove the lower bound, because, if  $\gamma \geq 2\beta$  the proof for this case coincides with proof of the lower bound in Theorem 1, where one needs to take  $f_0(\cdot) \equiv 0$  instead of  $f_0(\cdot) = f_0(\cdot, \vartheta_0)$ . In the case  $\beta < \gamma < 2\beta$ ,

the function  $\varphi_\varepsilon(\alpha_\varepsilon)$  coincides with  $\varphi_\varepsilon(\gamma)$ , being MRC on the hypothesis set  $\Sigma_0$ . Then, required lower bound follows from Proposition 1.

Theorem is proved.

6.2.2. *Proof of Lemma 5.* By definition

$$\begin{aligned} \hat{d}_\varepsilon &= B_\varepsilon^2(f) + 2\varepsilon \left( \langle b_\varepsilon(\cdot, f) - \tilde{b}_\varepsilon(\cdot, f) \rangle, \left( \sqrt{N_\varepsilon} \eta_\varepsilon(\cdot) - \sqrt{\tilde{N}_\varepsilon} \tilde{\eta}_\varepsilon(\cdot) \right) \right) \\ &+ \varepsilon^2 N_\varepsilon \|\eta_\varepsilon(\cdot)\|_2^2 - 2\varepsilon^2 \sqrt{N_\varepsilon \tilde{N}_\varepsilon} \langle \eta_\varepsilon(\cdot), \tilde{\eta}_\varepsilon(\cdot) \rangle \\ &+ \varepsilon^2 \tilde{N}_\varepsilon \|\tilde{\eta}_\varepsilon(\cdot)\|_2^2 + \varepsilon^2 (l+1) (\tilde{N}_\varepsilon - N_\varepsilon). \end{aligned}$$

Hence, in order to prove the statement of the lemma it is enough to show that

$$\langle b_\varepsilon(\cdot, f), \tilde{\eta}_\varepsilon(\cdot) \rangle = \langle \tilde{b}_\varepsilon(\cdot, f), \tilde{\eta}_\varepsilon(\cdot) \rangle; \quad (6.81)$$

$$\sqrt{N_\varepsilon} \langle \eta_\varepsilon(\cdot), \tilde{\eta}_\varepsilon(\cdot) \rangle = \sqrt{\tilde{N}_\varepsilon} \|\tilde{\eta}_\varepsilon(\cdot)\|_2^2. \quad (6.82)$$

Let us show, for example, (6.79). (6.78) is obtained by the same calculations.

Fix some  $k = \overline{1, \tilde{N}_\varepsilon}$  and consider

$$Z_k = \sqrt{N_\varepsilon} \int_{\tilde{\Delta}_k} \eta_\varepsilon(t) \tilde{\eta}_\varepsilon(t) \mathbf{d}t = \sqrt{N_\varepsilon} \sum_{i=kK_\varepsilon}^{(k+1)K_\varepsilon} \int_{\Delta_i} \eta_\varepsilon(t) \tilde{\eta}_\varepsilon(t) \mathbf{d}t.$$

Denote by  $V_i$ ,  $i = \overline{kK_\varepsilon, (k+1)K_\varepsilon}$ ,  $(l+1) \times (l+1)$ -matrix, which is defined as follows.

$$V_i = \int_{\Delta_i} d^{(i)}(t) \left( \tilde{d}^{(k)}(t) \right)' \mathbf{d}t.$$

Then

$$Z_k = \sqrt{N_\varepsilon} \sum_{i=kK_\varepsilon}^{(k+1)K_\varepsilon} \left( \xi^{(i)}(\varepsilon) \right)' \Lambda_l^{-1} V_i \Lambda_l^{-1} \left( \tilde{\xi}^{(k)}(\varepsilon) \right).$$

Let  $v_i(p, q)$ ,  $p, q = \overline{1, l+1}$ , is  $(p, q)$ -element of the matrix  $V_i$ . Direct calculations show that

$$v_i(p, q) = \tilde{N}_\varepsilon^{q-1} \sum_{r=0}^{q-1} \binom{q-1}{r} (t_{i-1} - \tilde{t}_{k-1})^{q-1-r} N_\varepsilon^{-(r+1)} \frac{1}{p+r}.$$

Let  $u_i(p, q)$ ,  $p, q = \overline{1, l+1}$ , is  $(p, q)$ -element of the matrix  $U_i = \Lambda_l^{-1} V_i$ . Again the simplest algebra yields

$$u_i(p, q) = \begin{cases} \tilde{N}_\varepsilon^{q-1} \binom{q-1}{p-1} (t_{i-1} - \tilde{t}_{k-1})^{q-p} N_\varepsilon^{-p} & q \geq p; \\ 0 & q < p. \end{cases}$$

Introduce the vector  $(\tau^{(i)})' = (\tau_1^{(i)}, \dots, \tau_{l+1}^{(i)}) = (\xi^{(i)}(\varepsilon))' U_i$ . Then  $\forall j = \overline{1, l+1}$  we have

$$\begin{aligned} \tau_j^{(i)} &= \sum_{p=1}^{l+1} \xi_p^{(i)}(\varepsilon) u_i(p, j) = \sum_{p=1}^j \xi_p^{(i)}(\varepsilon) u_i(p, j) \\ &= \tilde{N}_\varepsilon^{j-1} N_\varepsilon^{-\frac{1}{2}} \int_{\Delta_i} \mathbf{d}w(t) \left( \sum_{p=1}^j \binom{j-1}{p-1} (t - t_{i-1})^{p-1} (t_{i-1} - \tilde{t}_{k-1})^{j-p} \right) \\ &= \tilde{N}_\varepsilon^{j-1} N_\varepsilon^{-\frac{1}{2}} \int_{\Delta_i} (t - \tilde{t}_{k-1})^{j-1} \mathbf{d}w(t). \end{aligned}$$

Hence, we have

$$\bar{\tau}^{(k)} \triangleq \sum_{i=kK_\varepsilon}^{(k+1)K_\varepsilon} \tau^{(i)} = \tilde{N}_\varepsilon^{-\frac{1}{2}} N_\varepsilon^{-\frac{1}{2}} \tilde{\xi}^{(k)}(\varepsilon).$$

By definition

$$\begin{aligned} Z_k &= \sqrt{N_\varepsilon} (\bar{\tau}^{(k)})' \Lambda_l^{-1} \tilde{\xi}^{(k)}(\varepsilon) = \tilde{N}_\varepsilon^{-\frac{1}{2}} (\tilde{\xi}^{(k)}(\varepsilon))' \Lambda_l^{-1} \tilde{\xi}^{(k)}(\varepsilon) \\ &= \tilde{N}_\varepsilon^{-\frac{1}{2}} \|\tilde{\eta}_\varepsilon^{(k)}\|^2. \end{aligned}$$

It remains to note that

$$\begin{aligned} N_\varepsilon^{-\frac{1}{2}} \langle \eta_\varepsilon(\cdot), \tilde{\eta}_\varepsilon(\cdot) \rangle &= \sum_{k=1}^{\tilde{N}_\varepsilon} Z_k = \tilde{N}_\varepsilon^{-\frac{1}{2}} \sum_{k=1}^{\tilde{N}_\varepsilon} \|\tilde{\eta}_\varepsilon^{(k)}\|^2 \\ &= \tilde{N}_\varepsilon^{\frac{1}{2}} \|\tilde{\eta}_\varepsilon(\cdot)\|_2^2. \end{aligned}$$

Lemma is proved.

## 7. Appendix. Proofs of Propositions 1-3

**7.1. Proof of Proposition 1.** Fix an arbitrary  $\alpha_{(\cdot)}$  and let  $f_\varepsilon^*(\cdot)$  be some  $\alpha_{(\cdot)}$ -adaptive estimator. Introduce the following estimator

$$\tilde{f}_\varepsilon(\cdot) = \begin{cases} f_\varepsilon^*(\cdot) & \hat{\varrho}_\varepsilon = \varphi_\varepsilon(\alpha_{(\cdot)}); \\ \bar{f}_\varepsilon(\cdot) & \hat{\varrho}_\varepsilon = \varphi_\varepsilon(\Sigma). \end{cases}$$

Fix some  $\delta > 0$ . Since  $\hat{\varrho}_\varepsilon$  is optimal r.n.f. and  $\Sigma_0 \subset \Sigma$ , for all small enough  $\varepsilon > 0$  one has

$$\begin{aligned} M^* + \delta &\geq \sup_{f(\cdot) \in \Sigma_0} \mathbf{E}_f \left( \hat{\varrho}_\varepsilon^{-1} \|f_\varepsilon^*(\cdot) - f(\cdot)\|_p \right)^q \\ &\geq \sup_{f(\cdot) \in \Sigma_0} \mathbf{E}_f^{(\varepsilon)} \left( \varphi_\varepsilon^{-1}(\alpha_\varepsilon) \|f_\varepsilon^*(\cdot) - f(\cdot)\|_p \right)^q I\{\hat{\varrho}_\varepsilon = \varphi_\varepsilon(\alpha_{(\cdot)})\} \\ &= \sup_{f(\cdot) \in \Sigma_0} \mathbf{E}_f^{(\varepsilon)} \left( \varphi_\varepsilon^{-1}(\alpha_\varepsilon) \|\tilde{f}_\varepsilon(\cdot) - f(\cdot)\|_p \right)^q I\{\hat{\varrho}_\varepsilon = \varphi_\varepsilon(\alpha_{(\cdot)})\} \\ &= \pi_\varepsilon \sup_{f(\cdot) \in \Sigma_0} \mathbf{E}_f^{(\varepsilon)} \left( \varphi_\varepsilon^{-1}(\Sigma_0) \|\bar{f}_\varepsilon(\cdot) - f(\cdot)\|_p \right)^q \\ &= \pi_\varepsilon \sup_{f(\cdot) \in \Sigma_0} \mathbf{E}_f^{(\varepsilon)} \left( \varphi_\varepsilon^{-1}(\Sigma_0) \|\bar{f}_\varepsilon(\cdot) - f(\cdot)\|_p \right)^q I\{\hat{\varrho}_\varepsilon = \varphi_\varepsilon(\Sigma)\}, \quad (7.1) \end{aligned}$$

where  $\pi_\varepsilon = (\varphi_\varepsilon(\Sigma_0)\varphi_\varepsilon^{-1}(\alpha_\varepsilon))^q$ . By Cauchy-Schwartz inequality, for all small enough  $\varepsilon > 0$  we get

$$\begin{aligned} & \mathbf{E}_f^{(\varepsilon)} \left( \varphi_\varepsilon^{-1}(\Sigma_0) \|\bar{f}_\varepsilon(\cdot) - f(\cdot)\|_p \right)^q I\{\hat{\varrho}_\varepsilon = \varphi_\varepsilon(\Sigma)\}, \\ & \leq \left\{ \mathbf{E}_f^{(\varepsilon)} \left( \varphi_\varepsilon^{-1}(\Sigma_0) \|\bar{f}_\varepsilon(\cdot) - f(\cdot)\|_p \right)^{q_1} \right\}^{\frac{q}{q_1}} \left\{ P_f^{(\varepsilon)}\{\hat{\varrho}_\varepsilon = \varphi_\varepsilon(\Sigma)\} \right\}^{\frac{q_1 - q}{q_1}} \\ & \leq (R + \delta)^{\frac{q}{q_1}} \left( \sup_{f(\cdot) \in \Sigma_0} P_f^{(\varepsilon)}\{\hat{\varrho}_\varepsilon = \varphi_\varepsilon(\Sigma)\} \right)^{\frac{q_1 - q}{q_1}} \\ & \leq (R + \delta)^{\frac{q}{q_1}} (\alpha_0 + \delta)^{\frac{q_1 - q}{q_1}}. \end{aligned} \quad (7.2)$$

Here we used that  $\hat{\varrho}_\varepsilon \in \Omega_\varepsilon$  and the assumption of the proposition. From (7.2), continuing (7.1), we have

$$\begin{aligned} M^* + \delta & \geq \pi_\varepsilon \inf_{f_\varepsilon} \sup_{f(\cdot) \in \Sigma_0} R(f_\varepsilon, \Sigma_0, \varphi_\varepsilon(\Sigma_0)) \\ & \quad - (R + \delta) \frac{q}{q_1} (\alpha_0 + \delta)^{\frac{q_1 - q}{q_1}}. \end{aligned}$$

Hence

$$\liminf_{\varepsilon \rightarrow 0} \pi_\varepsilon^{-1} \geq \left( (M^* + \delta)^{-1} \left( l_0 - (R + \delta) \frac{q}{q_1} (\alpha_0 + \delta)^{\frac{q_1 - q}{q_1}} \right) \right)^{\frac{1}{q}}.$$

Tending  $\delta$  to zero we arrive at the statement of the proposition.

Proposition is proved.

7.1.1. *Proof of Proposition 2.* Since an estimated function is supposed to be bounded by some constant  $L$ , without loss of generality one can assume that  $\|\hat{f}_\varepsilon(\cdot)\|_p \leq 3L$ . As it has been already mention, each  $\alpha_{(\cdot)}$ -adaptive estimator is a.e. estimator on the set  $\Sigma$  (see Remark 2). Therefore, in order to establish the statement of the proposition, it is enough to show that

$$\limsup_{\varepsilon \rightarrow 0} R_\varepsilon(f_\varepsilon^*, \Sigma_0, \varphi_\varepsilon(\Sigma_0)) \leq \infty.$$

Indeed, due to the representation (3.1)

$$\begin{aligned} & R_\varepsilon(f_\varepsilon^*, \Sigma_0, \varphi_\varepsilon(\Sigma_0)) \\ & \leq \sup_{f(\cdot) \in \Sigma_0} \mathbf{E}_f^{(\varepsilon)} \left( \varphi_\varepsilon^{-1}(\Sigma_0) \|\hat{f}_\varepsilon^{(0)}(\cdot) - f(\cdot)\|_p \right)^q I\{\hat{\varrho}_\varepsilon = \varphi_\varepsilon(\alpha_\varepsilon)\} \\ & \quad + \sup_{f(\cdot) \in \Sigma_0} \mathbf{E}_f^{(\varepsilon)} \left( \varphi_\varepsilon^{-1}(\Sigma_0) \|\hat{f}_\varepsilon(\cdot) - f(\cdot)\|_p \right)^q I\{\hat{\varrho}_\varepsilon = \varphi_\varepsilon(\Sigma)\} \\ & \leq R_\varepsilon(\hat{f}_\varepsilon^{(0)}, \Sigma_0, \varphi_\varepsilon(\Sigma_0)) + (4L)^q \varphi_\varepsilon^{-q}(\Sigma_0) \sup_{f(\cdot) \in \Sigma_0} \mathbf{P}_f^{(\varepsilon)}\{\hat{\varrho}_\varepsilon = \varphi_\varepsilon(\Sigma)\}. \end{aligned} \quad (7.3)$$

Since  $\hat{f}_\varepsilon^{(0)}(\cdot)$  is a.e. estimator on the set  $\Sigma_0$

$$R_\varepsilon(\hat{f}_\varepsilon^{(0)}, \Sigma_0, \varphi_\varepsilon(\Sigma_0)) \leq \infty.$$

Since  $\hat{\varrho}_\varepsilon \in \Omega_\varepsilon$

$$\limsup_{\varepsilon \rightarrow 0} \alpha_\varepsilon^{-1} \mathbf{P}_f^{(\varepsilon)}\{\hat{\varrho}_\varepsilon = \varphi_\varepsilon(\Sigma)\} \leq 1.$$

Moreover, by assumption of the proposition  $\alpha_\varepsilon \varphi_\varepsilon^{-q}(\Sigma_0) = \mathcal{O}(1)$ . From these three facts and (7.3) we obtain the statement of the proposition.

Proposition is proved.

7.1.2. *Proof of Proposition 3.* Putting  $H_\gamma = \left(\frac{M^*}{\gamma}\right)^{\frac{1}{q}}$ , we have from the definition of the optimal r.n.f. and from the representation (3.1)

$$\begin{aligned} M^* &\geq \limsup_{\varepsilon \rightarrow 0} R(f_\varepsilon^*, \Sigma, \hat{\varrho}_\varepsilon) \\ &\geq \limsup_{\varepsilon \rightarrow 0} \sup_{f(\cdot) \in \Sigma_\varepsilon(H_\gamma)} \mathbf{E}_f^{(\varepsilon)} \left( \varphi_\varepsilon^{-1}(\alpha_\varepsilon) \|\hat{f}_\varepsilon^{(0)}(\cdot) - f(\cdot)\|_p \right)^q I\{\hat{\varrho}_\varepsilon = \varphi_\varepsilon(\alpha_\varepsilon)\}. \end{aligned} \quad (7.4)$$

Note that  $\|\hat{f}_\varepsilon^{(0)}(\cdot) - f(\cdot)\|_p \geq H_\gamma \varphi_\varepsilon(\alpha_\varepsilon)$  for all  $f(\cdot) \in \Sigma_\varepsilon(H_\gamma)$ , that follows from  $\hat{f}_\varepsilon^{(0)}(\cdot) \in \Sigma_0$  and from the definition of the set  $\Sigma_\varepsilon(H_\gamma)$ . Thus, continuing (7.4),

$$\begin{aligned} M^* &\geq H_\gamma^q \limsup_{\varepsilon \rightarrow 0} \sup_{f(\cdot) \in \Sigma_\varepsilon(H_\gamma)} \mathbf{P}_f^{(\varepsilon)}\{\hat{\varrho}_\varepsilon = \varphi_\varepsilon(\alpha_\varepsilon)\}. \\ &= \frac{M^*}{\gamma} \limsup_{\varepsilon \rightarrow 0} \sup_{f(\cdot) \in \Sigma_\varepsilon(H_\gamma)} \mathbf{P}_f^{(\varepsilon)}\{\hat{\varrho}_\varepsilon = \varphi_\varepsilon(\alpha_\varepsilon)\}. \end{aligned}$$

Proposition is proved.

## REFERENCES

- [1] Barron, A., Birge, L. and Massart, P. (1995). Risk bounds for model selection via penalization. *Preprint*, University Paris-Sud, Orsay.
- [2] Delyon, B. and Juditski, A. (1994). Wavelet estimators, global error measures, revisited. *Technical Report*, IRISA, Rennes.
- [3] Donoho, D.L. and Johnstone, I.M. (1992a). Ideal spatial adaptation by wavelet shrinkage *Biometrika* **81** 425–455.
- [4] Donoho, D.L., Johnstone, I.M., Kerkyacharian, G. and Picard, D. (1994). Wavelet shrinkage: asymptopia? *J. Royal Statist. Soc., Ser. B*, **57**, 301–369.
- [5] Efroimovich, S.Yu. and Pinsker, M.S. (1984). Adaptive algorithms for nonparametric filtering. *Automat. Remote Control* **11**, 54–60.
- [6] Efroimovich, S.Yu. (1985). Nonparametric estimation of a density of unknown smoothness. *Theory Probab. Appl.* **30**, 524–534.
- [7] Goldenshluger, A. and Nemirovski, A. (1994) On spatial adaptive estimation of nonparametric regression. *Technical report* **5**, Technion, Israel.
- [8] Golubev, G.K. (1990). Quasilinear estimates of a signal in  $L_2$ . *Problems Inform. Transmission* **26**, 15–20.
- [9] Härdle, W. and Marron, J.S. (1985). Optimal bandwidth selection in nonparametric regression function estimation. *Ann. Statist.* **12** 1466–1481.
- [10] Hall, P., Kerkyacharian, G. and Picard, D. (1995). On the minimax optimality of block thresholded wavelet estimators. Tentatively accepted to the *Statistica Sinica*.
- [11] Hall, P., Kerkyacharian, G. and Picard, D. (1996a). Block threshold rules for curve estimation using kernel and wavelet methods. Submitted to the *Annals of Statistics*.
- [12] Hall, P., Kerkyacharian, G. and Picard, D. (1996b). Note on the wavelet oracle. Submitted to the *Annals of Statistics*.
- [13] Ibragimov, I.A. and Khasminskii, R.Z. (1980). Estimates of a signal, its derivatives and point of maximum for Gaussian observation. *Theory Prob. Appl.* **25**, N.4, 703–716.
- [14] Ibragimov, I.A. and Khasminskii, R.Z. (1981). *Statistical Estimation: Asymptotic Theory* Springer, New York.

- [15] Ingster, Yu.I. (1993). Asymptotically minimax hypothesis testing for nonparametric alternatives. I-III. *Math. Methods of Statist.* **2** (1993) 85 – 114, **3** (1993) 171 – 189, **4** (1993) 249 – 268.
- [16] Jennrich, R.I.(1969). Asymptotic properties of non-linear least squares estimators. *Annals of Math. Statist.* **40**, N. 2, 633-643.
- [17] Korostelev,A.P. and Tsybakov,A.B.(1993). *Minimax Theory of Image Reconstruction*. Springer, New York.
- [18] Lepski,O.V.(1990). One problem of adaptive estimation in Gaussian white noise. *Theory Probab. Appl.* **35** N.3 459-470.
- [19] Lepski,O.V.(1991). Asymptotic minimax adaptive estimation. 1. Upper bounds. *Theory Probab. Appl.* **36** No.4, 645-659.
- [20] Lepski,O.V.(1992a). Asymptotic minimax adaptive estimation. 2. Statistical model without optimal adaptation. Adaptive estimators *Theory Probab. Appl.* **37** N.3, 468-481.
- [21] Lepski,O.V.(1992b). On problems of adaptive estimation in white Gaussian noise. *Advances in sov. math.*, R.Z.Khasminskii ed., **12**, 87-106.
- [22] Lepski, O.V. and Spokoiny, V.G. (1995a). Local adaptation to inhomogeneous smoothness: resolution level. *Math. Meth. Statist.* **4**, N. 3, 239-258.
- [23] Lepski, O.V., Mammen, E. and Spokoiny, V.G. (1995). Ideal spatial adaptation to inhomogeneous smoothness: an approach based on kernel estimates with variable bandwidth selection. *Ann. Statist.*, tentatively accepted.
- [24] Lepski, O.V. and Spokoiny, V.G. (1995b). Optimal Pointwise Adaptive Methods in Nonparametric Estimation. *Ann. Statist.*, tentatively accepted.
- [25] Low, M. (1996). On nonparametric confidence intervals. *Ann. Statist.*, tentatively accepted.
- [26] Linton,O. and Nielsen,J.P. (1995) A kernel methods of estimating structured nonparametric regression based on marginal integration. *Biometrika* **82**, 93-101.
- [27] Neumann,M.H. and von Sachs,R.(1995). Wavelet threshholding in anisotropic function classes and application to the adaptive estimation of evolutionary spectra. *Preprint WIAS*, Berlin.
- [28] Nikolskii,S.M.(1975). *Approximation of Function of Several Variables and Imbedding Theorems*. Springer, Berlin.
- [29] Nussbaum, M. (1986). On nonparametric estimation of a regression function, being smooth on a domain in  $R^k$ . *Theory Probab. Appl.* **31**, N. 2, 118-125.
- [30] Petrov,V.V.(1975) *Sums of Independent Random Variables*. Springer, New York.
- [31] Polyak,B.T and Tsybakov,A.B.(1990). Asymptotical optimality of the  $C_p$ -test for the orthogonal series estimators of regression. *Theory Probab. Appl.* **35**, N. 2, 293-306.
- [32] Speckman,P.(1988). Kernel smoothing in partial linear models. *Journal of the Royal Statistical Society, Series B* **50**, 413-446.
- [33] Stone,C.J.(1985). Additive regression and other nonparametric models. *Ann. Stat.* **13**, 689-705.