

# Adaptive minimax estimation of a fractional derivative

Farida Enikeeva<sup>a,b</sup>

<sup>a</sup>EURANDOM, P.O. Box 513, 5600 MB Eindhoven, The Netherlands

<sup>b</sup>Institute for Information Transmission Problems, Bolshoy Karetnyi 19, GSP-4, Moscow 127994, Russia

Received 29 March 2005; received in revised form 27 February 2006; accepted 1 March 2006

Available online 18 April 2006

## Abstract

This paper considers a problem of adaptation in estimating a fractional antiderivative of an unknown drift density from observations in Gaussian white noise. This problem is closely related to the Wicksell problem. Under the assumption that the drift density belongs to a Sobolev class with unknown smoothness, an adaptive estimator is constructed.

© 2006 Elsevier B.V. All rights reserved.

*Keywords:* Adaptive estimation; Oracle inequality; White noise model; The Wicksell problem

## 1. Introduction

We observe noisy data

$$X_k = \theta_k + \varepsilon \zeta_k, \quad k = 1, 2, \dots, \quad (1)$$

where  $\zeta_k$  are i.i.d.  $\mathcal{N}(0, 1)$ , and the parameter  $\varepsilon > 0$  is assumed to be known. Our goal is to recover a vector  $v(\theta) = (v_1(\theta), v_2(\theta), \dots)$ , with components  $v_k(\theta) = \theta_k / \sqrt{k}$ , such that  $v(\theta) \in \ell_2$ .

The problem of estimating  $v(\theta)$  was recently considered by Golubev and Enikeeva (2001). There, it is assumed that the vector  $\theta = (\theta_1, \theta_2, \dots)$  belongs to a certain ellipsoid  $\Theta$ :

$$\theta \in \Theta = \left\{ \theta : \sum_{k=1}^{\infty} a_k^2 \theta_k^2 \leq 1 \right\}, \quad (2)$$

with fixed coefficients  $\{a_k\}$ . For example, if  $\Theta = \Theta_\beta(P)$  is a Sobolev ellipsoid with the smoothness parameter  $\beta$  and radius  $P$ , then  $a_k = (\pi k)^\beta / \sqrt{P}$ . Under assumption (2), the authors follow the classical approach of Pinsker (1980) to obtain an asymptotically minimax estimator of  $v(\theta)$ . Unfortunately,  $a_k$ , the parameters of the ellipsoid, often cannot be completely specified a priori. Moreover, the estimator in Golubev and Enikeeva, 2001 depends on an implicitly given smoothness parameter. Therefore, there arises the problem of adaptive estimation. In adaptive estimation, one usually has a list of models, for example, a family of Sobolev ellipsoids  $\Theta_\beta(P)$  where  $P$  is fixed, the parameter  $\beta$  belongs to some set  $\mathcal{B}$ , but otherwise is unknown. It is then desirable

*E-mail addresses:* [enikeeva@eurandom.tue.nl](mailto:enikeeva@eurandom.tue.nl), [enikeeva@iitp.ru](mailto:enikeeva@iitp.ru).

to construct an estimator that depends only on the observations  $X_1, X_2, \dots$  and is asymptotically minimax for any  $\Theta_\beta, \beta \in \mathcal{B}$ . Such an estimator is called an *adaptive estimator*.

To motivate our investigation, consider the stochastic differential equation

$$dx(t) = g(t) dt + \varepsilon dw(t), \quad t \in [0, 1], \quad x(0) = 0, \tag{3}$$

where  $w(t)$  is the standard Wiener process,  $\varepsilon > 0$  is a small parameter, and the drift density  $g(t)$  is an unknown periodic function. We can consider the observations (3) in the domain of their Fourier coefficients:

$$\tilde{X}_k = \tilde{\theta}_k + \varepsilon \tilde{\xi}_k, \quad k = \pm 1, \pm 2, \dots, \tag{4}$$

where

$$\tilde{X}_k = \int_0^1 \phi_k(t) dx(t), \quad \tilde{\theta}_k = \int_0^1 \phi_k(t) g(t) dt$$

and  $\tilde{\xi}_k = \int_0^1 \phi_k(t) dw(t)$  are i.i.d.  $\mathcal{N}(0, 1)$ ;  $\{\phi_k\}$  is the trigonometric basis of  $\mathbb{L}_2(0, 1)$ .

It is well-known that the derivative of order  $\alpha \in \mathbb{R}$  of the function  $g(t)$  can be defined by the following formula (Zygmund, 1968):

$$g^{(\alpha)}(t) = \sum_{k=-\infty}^{\infty} \tilde{\theta}_k \phi_k(t) (2\pi i k)^\alpha,$$

and, consequently,

$$g^{(-1/2)}(t) = \sum_{k=-\infty}^{\infty} \frac{\tilde{\theta}_k}{\sqrt{k}} \phi_k(t) (2\pi i)^{-1/2}.$$

The derivative of negative order is called an antiderivative. Thus, the problem of estimating  $v(\theta)$  from the observations (1) is similar to the problem of recovering the fractional antiderivative of order  $-\frac{1}{2}$  from the observations (3).

The latter problem is, in turn, closely related to the Wicksell problem (Wicksell, 1925), formulated as follows: a number of spheres are embedded in an opaque medium. Let their radii be i.i.d. with an unknown distribution function  $F(x)$ . Since the medium is opaque, we cannot observe the radii of spheres directly. Instead, we intersect the medium by a plane and observe resulting circular cross-sections. Let  $Y_1, \dots, Y_n$  be the squared radii of the cross-sectional circles. The problem is to estimate the distribution function  $F(x)$  from these observations. Under some reasonable assumptions, it can be seen (Stoyan et al., 1995) that the random variables  $Y_i$  are i.i.d.; denote their distribution function by  $G(y)$ . The relation between  $F$  and  $G$  is well-known:

$$1 - G(y) = \int_y^\infty \sqrt{x - y} dF(x) \left( \int_0^\infty \sqrt{x} dF(x) \right)^{-1}.$$

If  $F$  is a Lipschitz function, this equation can be solved

$$F(x) = 1 - \frac{G^{(1/2)}(x)}{G^{(1/2)}(0)} \equiv 1 - \frac{g^{(-1/2)}(x)}{g^{(-1/2)}(0)},$$

where  $g$  is the density of  $G$ . We refer the reader to the paper of Golubev and Levit (1998) for a derivation of these formulas. Thus, in order to construct an estimator in the Wicksell problem we have to estimate the fractional antiderivatives of the density  $g$  at zero and on  $\mathbb{R}^+$ . Obviously, the Wicksell problem does not coincide with the problem of estimating the fractional antiderivative of the drift density in Gaussian white noise model. However, they are related closely. Suppose that  $g$  belongs to a small neighborhood of the uniform density on  $[0, 1]$ . Then, on certain conditions, for  $\varepsilon = n^{-1/2}$  in (3) the corresponding statistical experiments are asymptotically equivalent in the Le Cam sense (Nussbaum, 1996, p. 2409).

In this paper we construct an adaptive asymptotically minimax estimator of the vector  $v(\theta)$  under the assumption that  $\theta$  belongs to the Sobolev class  $\Theta_\beta$  with unknown smoothness  $\beta$ . In Section 2 we formulate the main result and describe a method of estimation and adaptation. Section 3 contains some auxiliary lemmas. The proof of the main result and concluding remarks can be found in Section 4.

## 2. Adaptive estimation

We observe the data

$$X_k = \theta_k + \varepsilon \xi_k, \quad k = 1, 2, \dots \tag{5}$$

We would like to construct an adaptive estimator of the unknown vector  $v(\theta) = (v_1, v_2, \dots)^T$  with components  $v_k = \theta_k/\sqrt{k}$  from these observations with the only assumption  $v(\theta) \in \ell_2$ . Denote for brevity the vector  $(X_1, X_2, \dots)^T$  by  $X$ .

Let  $\hat{v}(X) = (\hat{v}_1(X), \hat{v}_2(X), \dots)^T$  be an estimator of  $v(\theta)$ . Define the mean-square risk of  $\hat{v}$ :

$$\mathbf{E}_\theta \|\hat{v}(X) - v(\theta)\|^2 = \mathbf{E}_\theta \sum_{k=1}^\infty |\hat{v}_k - v_k|^2,$$

where  $\mathbf{E}_\theta$  is the expectation with respect to the measure corresponding to the distribution of  $X$ .

We will look for an adaptive estimator of  $v(\theta)$  in the class  $\mathcal{P}$  of projection estimators:

$$\mathcal{P} = \left\{ \hat{v}(W, X) : \hat{v}_k(W, X) = \lambda_k(W) \frac{X_k}{\sqrt{k}} \right\},$$

where

$$\lambda_k(W) = \begin{cases} 1, & k \leq W, \\ 0 & \text{otherwise.} \end{cases}$$

The integer parameter  $W$  is called *the bandwidth* of the projection estimate. We denote the corresponding projection estimator by  $\hat{v}(W)$  and its mean-square risk by  $R_\varepsilon(W, \theta)$ . Our aim is to find the best projection estimator of the vector  $\theta$ . It is easy to calculate the risk of  $\hat{v}(W)$ :

$$R_\varepsilon(W, \theta) = \mathbf{E}_\theta \|\hat{v}(W) - v(\theta)\|^2 = \varepsilon^2 \sum_{k=1}^W \frac{1}{k} + \sum_{k=W+1}^\infty \frac{\theta_k^2}{k}. \tag{6}$$

The choice of the class of projection estimators for adaptation is suggested by the minimax approach. Let us return for a moment to the problem where prior information is available. Suppose that  $\theta$  belongs to the Sobolev ellipsoid  $\Theta_\beta$ :

$$\sum_{k=1}^\infty a_k^2 \theta_k^2 \leq 1, \quad a_k^2 = (\pi k)^{2\beta} / P.$$

Taking into account this assumption, we can bound the risk (6) of the projection estimator  $\hat{v}(W)$  from above

$$\begin{aligned} R_\varepsilon(W, \theta) &= \varepsilon^2 \sum_{k=1}^W \frac{1}{k} + \sum_{k=W+1}^\infty \frac{\theta_k^2}{k} \leq \varepsilon^2 (\log W + \gamma + o(1)) + \sum_{k=W+1}^\infty \theta_k^2 a_k^2 \cdot \frac{1}{ka_k^2} \\ &\leq \varepsilon^2 (\log W + \gamma + o(1)) + \sup_{k > W} \frac{P}{\pi^{2\beta}} \cdot k^{-2\beta-1} \\ &\leq \varepsilon^2 (\log W + \gamma + o(1)) + \frac{P}{\pi^{2\beta}} W^{-2\beta-1}. \end{aligned}$$

Minimizing the last expression with respect to  $W$  we get

$$W_\beta^* = \left( \frac{P(2\beta + 1)}{\pi^{2\beta} \varepsilon^2} \right)^{1/2\beta+1}. \tag{7}$$

Thus, an upper bound on the mean square risk is

$$\sup_{\theta \in \Theta_\beta} R_\varepsilon(W_\beta^*, \theta) \leq \frac{\varepsilon^2}{2\beta + 1} \log \frac{P(2\beta + 1)}{\varepsilon^2 \pi^{2\beta}} + \varepsilon^2 \left( \gamma + \frac{1}{2\beta + 1} \right) + o(\varepsilon^2).$$

From Golubev and Enikeeva (2001) we have a lower bound on the risk and, consequently, the asymptotically minimax risk of the second order in this case is

$$\inf_{\hat{v}} \sup_{\theta \in \Theta_\beta} R_\varepsilon(W, \theta) = \frac{\varepsilon^2}{2\beta + 1} \log \frac{P(2\beta + 1)}{\varepsilon^2 \pi^{2\beta}} + \varepsilon^2 \left( \gamma - \frac{2}{2\beta + 1} \right) + o(\varepsilon^2).$$

Thus, the projection estimator is asymptotically minimax on the Sobolev ellipsoid  $\Theta_\beta$ . Our goal is to find an adaptive minimax estimator in the class of projection estimators but with  $W$  data dependent.

An estimator  $\hat{v}$  of the vector  $v(\theta)$  is *exactly adaptive in minimax sense* on the family of classes  $\Theta_\beta, \beta \in \mathcal{B}$  if

$$\lim_{\varepsilon \rightarrow 0} \frac{\sup_{\theta \in \Theta_\beta} \mathbf{E}_\theta \|\hat{v} - v(\theta)\|^2}{\inf_{\tilde{v}} \sup_{\theta \in \Theta_\beta} \mathbf{E}_\theta \|\tilde{v} - v(\theta)\|^2} = 1 \quad \forall \beta \in \mathcal{B}.$$

Let us return to the problem of adaptive choice of  $W$ . If  $\theta = (\theta_1, \theta_2, \dots)$  were known, then an optimal bandwidth could be found as the minimizer of the functional  $R_\varepsilon(W, \theta)$ :

$$W^{\text{oracle}} = \arg \min_W R_\varepsilon(W, \theta).$$

Indeed, we cannot do better without knowing  $\theta$ . We will call a map  $\theta \mapsto \hat{v}(W^{\text{oracle}})$  an *oracle* and the value

$$R_\varepsilon(W^{\text{oracle}}, \theta) = \min_W R_\varepsilon(W, \theta)$$

the *oracle risk*. Hereafter, we will also call the bandwidth  $W^{\text{oracle}}$  *oracle*. Of course,  $\hat{v}(W^{\text{oracle}})$  is not an estimator because it depends on  $\theta$  that we cannot know. However, we attempt to construct an estimator which will adapt to the oracle in the sense of imitating the oracle risk.

More precisely, an estimator  $\hat{v}(W)$  is called *adaptive to the oracle*  $W^{\text{oracle}}$  on the set  $\Theta$  if there exists a constant  $C < \infty$  such that

$$R_\varepsilon(W, \theta) \leq CR_\varepsilon(W^{\text{oracle}}, \theta) \tag{8}$$

for all  $\theta \in \Theta$  and  $0 < \varepsilon < 1$ .

An estimator  $\hat{v}(W)$  is *exactly adaptive to the oracle*  $W^{\text{oracle}}$  on the set  $\Theta$  if for all  $\theta \in \Theta$  we have

$$R_\varepsilon(W, \theta) \leq (1 + o(1))R_\varepsilon(W^{\text{oracle}}, \theta), \tag{9}$$

where  $o(1) \rightarrow 0$ , as  $\varepsilon \rightarrow 0$  uniformly in  $\theta \in \Theta$ .

Inequalities of the type (8), (9) are called *oracle inequalities*.

We would like to find an optimal bandwidth  $\hat{W}$  such that the risk of the corresponding estimator  $R_\varepsilon(\hat{W}, \theta)$  converges to the risk of the oracle, as  $\varepsilon \rightarrow 0$ . The general method to find such an estimator is based on the idea of unbiased risk estimation. This method goes back to the works of Mallows (1973) and Akaike (1973).

It is easy to see that  $X_k^2 - \varepsilon^2$  is an unbiased estimator of the parameter  $\theta_k^2$ :

$$\mathbf{E}_\theta(X_k^2 - \varepsilon^2) = \theta_k^2.$$

Thus, substituting  $\theta_k^2$  by this estimate in  $R_\varepsilon(W, \theta)$ , we arrive at an unbiased estimate of the risk

$$\begin{aligned} R_\varepsilon(W, \theta) &= \varepsilon^2 \sum_{k=1}^W \frac{1}{k} + \sum_{k=W+1}^\infty \frac{\theta_k^2}{k} \\ &= \varepsilon^2 \sum_{k=1}^W \frac{1}{k} + \|v(\theta)\|^2 - \sum_{k=1}^W \frac{\theta_k^2}{k} = 2\varepsilon^2 \sum_{k=1}^W \frac{1}{k} + \|v(\theta)\|^2 - \mathbf{E}_\theta \sum_{k=1}^W \frac{X_k^2}{k}. \end{aligned}$$

It follows that

$$R_\varepsilon(W, \theta) - \|v(\theta)\|^2 = 2\varepsilon^2 \sum_{k=1}^W \frac{1}{k} - \mathbf{E}_\theta \sum_{k=1}^W \frac{X_k^2}{k} = \mathbf{E}_\theta U(W, X),$$

where

$$U(W, X) = 2\varepsilon^2 \sum_{k=1}^W \frac{1}{k} - \sum_{k=1}^W \frac{X_k^2}{k}.$$

Therefore,  $U(W, X)$  is unbiased estimator of the risk  $R_\varepsilon(W, \theta)$  up to the constant  $\|v(\theta)\|$

$$R_\varepsilon(W, \theta) - \|v(\theta)\|^2 = \mathbf{E}_\theta U(W, X).$$

Now, to find an optimal  $W$  we minimize the functional  $U(W, X)$  in  $W$

$$\widehat{W} = \arg \min_{W \in \mathbb{N}} U(W, X). \tag{10}$$

We arrive at

**Theorem 1.** Let  $\widehat{W}$  be as in (10). For any  $\alpha \in (0, 1)$  the following oracle inequality holds:

$$R_\varepsilon(\widehat{W}, \theta) \leq \frac{1}{1 - \alpha} \min_{W \in \mathbb{N}} R_\varepsilon(W, \theta) + \varepsilon^2 C(\alpha), \tag{11}$$

for every  $v(\theta) \in \ell_2$  and for

$$C(\alpha) = \frac{1}{1 - \alpha} \left( \sqrt{\frac{2}{3}}\pi + \frac{2}{\alpha} \right).$$

We postpone the proof until Section 4.

**Remark 1.** It follows from the oracle inequality (11) that the estimator  $\widehat{v}(\widehat{W})$  is exactly adaptive to the oracle  $W^{\text{oracle}}$  for all  $v(\theta) \in \ell_2$ .

**Proof.** Indeed, take  $\alpha = (\log \log \varepsilon^{-2})^{-1}$ . Then we have for any  $v(\theta) \in \ell_2$

$$\begin{aligned} R_\varepsilon(\widehat{W}, \theta) &\leq (1 + (\log \log \varepsilon^{-2})^{-1}) R(W^{\text{oracle}}, \theta) + 2\varepsilon^2 \log \log \varepsilon^{-2} (1 + o(1)) \\ &\leq (1 + o(1)) R(W^{\text{oracle}}, \theta), \quad \varepsilon \rightarrow 0. \quad \square \end{aligned}$$

**Remark 2.** The constructed adaptive to the oracle estimator  $\widehat{v}(\widehat{W})$  is exactly adaptive in minimax sense on the family of Sobolev ellipsoids  $\{\Theta_\beta, \beta > \frac{1}{2}\}$ :

$$\lim_{\varepsilon \rightarrow 0} \frac{\sup_{\theta \in \Theta_\beta} \mathbf{E}_\theta \|\widehat{v}(\widehat{W}) - v(\theta)\|^2}{\inf_v \sup_{\theta \in \Theta_\beta} \mathbf{E}_\theta \|\widetilde{v} - v(\theta)\|^2} = 1, \quad \forall \beta > \frac{1}{2}.$$

**Proof.** Let  $\beta$  be fixed. From the oracle inequality it follows that

$$\sup_{\theta \in \Theta_\beta} \mathbf{E}_\theta \|\widehat{v}(\widehat{W}) - v(\theta)\|^2 \leq \frac{1}{1 - \alpha} \sup_{\theta \in \Theta_\beta} \mathbf{E}_\theta \|\widehat{v}(W^{\text{oracle}}) - v(\theta)\|^2 + \varepsilon^2 C(\alpha).$$

Then, for the optimal bandwidth  $W_\beta^*$  from (7),

$$\begin{aligned} \sup_{\theta \in \Theta_\beta} \mathbf{E}_\theta \|\widehat{v}(W^{\text{oracle}}) - v(\theta)\|^2 &\leq \sup_{\theta \in \Theta_\beta} \mathbf{E}_\theta \|\widehat{v}(W_\beta^*) - v(\theta)\|^2 \\ &\leq \frac{1}{2\beta + 1} \varepsilon^2 \log \varepsilon^{-2} (1 + o(1)), \quad \varepsilon \rightarrow 0. \end{aligned}$$

Thus, for any ellipsoid  $\Theta_\beta$ , and for a sequence  $\alpha = \alpha(\varepsilon) = (\log \log \varepsilon^{-2})^{-1}$ ,  $\varepsilon \rightarrow 0$ , we have

$$\sup_{\theta \in \Theta_\beta} \mathbf{E}_\theta \|\widehat{v}(\widehat{W}) - v(\theta)\|^2 \leq \frac{1}{2\beta + 1} \varepsilon^2 \log \varepsilon^{-2} (1 + o(1)).$$

As it was mentioned before, the lower bound of the minimax risk for Sobolev ellipsoids has the same form (see Golubev and Enikeeva, 2001):

$$\inf_{\hat{v}} \sup_{\theta \in \Theta_\beta} \mathbf{E}_\theta \|\hat{v} - v(\theta)\|^2 \geq \frac{1}{2\beta + 1} \varepsilon^2 \log \varepsilon^{-2} (1 + o(1)).$$

It follows that the estimator  $\hat{v}(\hat{W})$  is asymptotically minimax efficient for any ellipsoid  $\Theta_\beta$ :

$$\lim_{\varepsilon \rightarrow 0} \frac{\sup_{\theta \in \Theta_\beta} \mathbf{E}_\theta \|\hat{v}(\hat{W}) - v(\theta)\|^2}{\inf_{\hat{v}} \sup_{\theta \in \Theta_\beta} \mathbf{E}_\theta \|\hat{v} - v(\theta)\|^2} = 1, \quad \forall \beta > \frac{1}{2}.$$

This estimator is adaptive and does not depend on the smoothness parameter of the ellipsoid  $\Theta_\beta$ .  $\square$

Oracle inequalities in minimax adaptive constructions appeared in the works of Golubev and Nussbaum (see Golubev and Nussbaum, 1992 and references therein). We refer the reader to the paper of Kneip (1993) for an extensive bibliography on data-driven choice of smoothing parameters. More recent references are Donoho and Johnstone (1995), Birgé and Massart (2001), Cavalier et al. (2002), Tsybakov (2004).

### 3. Auxiliary tools

To prove the main result we need two auxiliary lemmas.

**Lemma 1.** *Let  $v$  be a positive integer random variable,  $\xi_k$  be i.i.d. standard Gaussian random variables. Then*

$$\mathbf{E} \sum_{k=1}^v \frac{\xi_k^2 - 1}{k} \leq \sqrt{\frac{2}{3}} \pi. \tag{12}$$

**Proof.** Let us note that

$$\mathbf{E} \sum_{k=1}^v \frac{\xi_k^2 - 1}{k} \leq \mathbf{E} \max_{m \in \mathbb{N}} \sum_{k=1}^m \frac{\xi_k^2 - 1}{k} = \lim_{N \rightarrow \infty} \mathbf{E} \left( \max_{1 \leq m \leq N} \sum_{k=1}^m \frac{\xi_k^2 - 1}{k} \right). \tag{13}$$

It is easy to see that the sequence  $|\sum_{k=1}^m \xi_k^2 - 1/k|$  is a non-negative submartingale bounded in  $\mathcal{L}_2$ , thus we can apply Doob’s  $\mathcal{L}_p$  inequality (Williams, 1991, p. 143) taking  $p = q = 2$ :

$$\left( \mathbf{E} \max_{1 \leq m \leq N} \left| \sum_{k=1}^m \frac{\xi_k^2 - 1}{k} \right|^2 \right)^{1/2} \leq 2 \left( \mathbf{E} \left| \sum_{k=1}^m \frac{\xi_k^2 - 1}{k} \right|^2 \right)^{1/2}.$$

Since  $\xi_k$  are standard Gaussian, we have

$$\mathbf{E} \left( \sum_{k=1}^N \frac{\xi_k^2 - 1}{k} \right)^2 = \mathbf{E} \sum_{k=1}^N \frac{(\xi_k^2 - 1)^2}{k^2} = 2 \sum_{k=1}^N \frac{1}{k^2}.$$

Therefore,

$$\mathbf{E} \max_{1 \leq m \leq N} \left| \sum_{k=1}^m \frac{\xi_k^2 - 1}{k} \right| \leq 4 \sum_{k=1}^N \frac{1}{k^2}.$$

Next, from the Jensen inequality we have

$$\mathbf{E} \max_{1 \leq m \leq N} \sum_{k=1}^m \frac{\xi_k^2 - 1}{k} \leq \left( \mathbf{E} \max_{1 \leq m \leq N} \left( \sum_{k=1}^m \frac{\xi_k^2 - 1}{k} \right)^2 \right)^{1/2} \leq 2 \left( \sum_{k=1}^N \frac{1}{k^2} \right)^{1/2}.$$

Applying this inequality to (13), we get

$$\lim_{N \rightarrow \infty} \mathbf{E} \left( \max_{1 \leq m \leq N} \sum_{k=1}^m \frac{\xi_k^2 - 1}{k} \right) \leq 2 \left( \sum_{k=1}^{\infty} \frac{1}{k^2} \right)^{1/2} = \sqrt{\frac{2}{3}} \pi.$$

Thus, the lemma follows.  $\square$

**Lemma 2.** Let  $v$  be a positive integer random variable,  $\xi_k$  be i.i.d. standard Gaussian random variables,  $v(\theta) \in \ell_2$ , and  $\alpha \in (0, 1)$ . Then

$$\mathbf{E}_\theta \left( 2\varepsilon \sum_{k=v+1}^{\infty} \frac{\theta_k \xi_k}{k} + \alpha \left( \varepsilon^2 \sum_{k=1}^v \frac{1}{k} + \sum_{k=v+1}^{\infty} \frac{\theta_k^2}{k} \right) \right) \geq -\frac{2\varepsilon^2}{\alpha}. \tag{14}$$

**Proof.** Note that

$$\sum_{k=v+1}^{\infty} \frac{\theta_k \xi_k}{k} = w \left( \sum_{k=v+1}^{\infty} \frac{\theta_k^2}{k^2} \right),$$

where  $w(t)$  is a standard Wiener process. Applying the following property of the Wiener process:

$$\mathbf{E} \min_{t \geq 0} \left\{ w(t) + \frac{\mu t}{2} \right\} \geq -\frac{1}{\mu}$$

we can bound the left-hand side of (14) from above. Set  $t_0 = \sum_{k=v+1}^{\infty} (\theta_k^2/k^2)$ . Then

$$\begin{aligned} & \mathbf{E}_\theta \left\{ 2\varepsilon \sum_{k=v+1}^{\infty} \frac{\theta_k \xi_k}{k} + \alpha \left( \varepsilon^2 \sum_{k=1}^v \frac{1}{k} + \sum_{k=v+1}^{\infty} \frac{\theta_k^2}{k} \right) \right\} \\ & \geq \mathbf{E} \{ 2\varepsilon w(t_0) + \alpha t_0 \} = 2\varepsilon \mathbf{E} \left\{ w(t_0) + \frac{\alpha}{2\varepsilon} t_0 \right\} \\ & \geq 2\varepsilon \mathbf{E} \min_{t \geq 0} \left\{ w(t) + \frac{\alpha}{2\varepsilon} t \right\} \geq -\frac{2\varepsilon^2}{\alpha}. \quad \square \end{aligned}$$

**4. Proof of the main result**

Now we can prove the main result.

**Proof.** It is easy to see that for any  $\alpha \in (0, 1)$

$$\begin{aligned} \mathbf{E}_\theta U(\widehat{W}, X) &= 2\varepsilon^2 \mathbf{E}_\theta \sum_{k=1}^{\widehat{W}} \frac{1}{k} - \mathbf{E}_\theta \sum_{k=1}^{\widehat{W}} \frac{\theta_k^2}{k} - 2\varepsilon \mathbf{E}_\theta \sum_{k=1}^{\widehat{W}} \frac{\theta_k \xi_k}{k} - \varepsilon^2 \mathbf{E}_\theta \sum_{k=1}^{\widehat{W}} \frac{\xi_k^2}{k} \\ &= (1 - \alpha) R_\varepsilon(\widehat{W}, \theta) - \|v(\theta)\|^2 \\ &\quad + 2\varepsilon \mathbf{E}_\theta \sum_{\widehat{W}+1}^{\infty} \frac{\theta_k \xi_k}{k} + \alpha \mathbf{E}_\theta \left( \varepsilon^2 \sum_{k=1}^{\widehat{W}} \frac{1}{k} + \sum_{k=\widehat{W}+1}^{\infty} \frac{\theta_k^2}{k} \right) - \varepsilon^2 \mathbf{E}_\theta \sum_{k=1}^{\widehat{W}} \frac{\xi_k^2 - 1}{k}. \end{aligned}$$

We can bound this equality from below using Lemmas 1 and 2:

$$\mathbf{E}_\theta U(\widehat{W}, X) \geq (1 - \alpha) R_\varepsilon(\widehat{W}, \theta) - \|\theta\|^2 - \frac{2\varepsilon^2}{\alpha} - \varepsilon^2 \sqrt{\frac{2}{3}} \pi. \tag{15}$$

Therefore, taking into account that for any  $W$

$$\mathbf{E}_\theta U(\widehat{W}, X) \leq \mathbf{E}_\theta U(W, X) \equiv R_\varepsilon(W, \theta) - \|\theta\|^2,$$

we can rewrite (15) as

$$R_\varepsilon(W, \theta) \geq (1 - \alpha)R_\varepsilon(\widehat{W}, \theta) - \frac{2\varepsilon^2}{\alpha} - \varepsilon^2 \sqrt{\frac{2}{3}}\pi,$$

and, consequently, for any  $W$

$$R_\varepsilon(\widehat{W}, \theta) \leq \frac{1}{1 - \alpha}R_\varepsilon(W, \theta) + \frac{\varepsilon^2}{1 - \alpha} \left( \sqrt{\frac{2}{3}}\pi + \frac{2}{\alpha} \right).$$

Thus, the theorem follows.  $\square$

**Concluding remarks.** We discussed the open question of adaptation in the Wicksell problem by considering a similar problem of adaptation in estimating the fractional antiderivative of the signal  $g$  in the white noise model. In the latter problem, we consider two cases: estimating the antiderivative  $g^{(-1/2)}$  on  $\mathbb{R}^+$  and at 0. These two cases are equivalent, correspondingly, to the problems of adaptive estimation of a vector  $v(\theta) = (\theta_1/\sqrt{1}, \theta_2/\sqrt{2}, \dots)$  and a linear functional  $L(\theta) = \sum_{k=1}^{\infty} \theta_k/\sqrt{k}$  from the observations (1). In the present paper we solved the first problem. In future work, we intend to treat the second case, using the method of adaptation for linear functionals recently proposed by Golubev (2004).

### Acknowledgments

The author thanks Youri Golubev for helpful advice and an anonymous referee for comments, which improved the presentation of the paper.

### References

- Akaike, H., 1973. Information theory and an extension of the maximum likelihood principle. In: Proceedings of the Second International Symposium Information Theory, Budapest, pp. 267–281.
- Birgé, L., Massart, P., 2001. Gaussian model selection. *J. Eur. Math. Soc.* 3, 203–268.
- Cavalier, L., Golubev, G.K., Picard, D., Tsybakov, A.B., 2002. Oracle inequalities for inverse problems. *Ann. Statist.* 30, 843–874.
- Donoho, D.L., Johnstone, I.M., 1995. Adapting to unknown smoothness via wavelet shrinkage. *J. Amer. Statist. Assoc.* 90, 1200–1224.
- Golubev, G., Levit, B., 1998. Asymptotically efficient estimation in the wicksell problem. *Ann. Statist.* 26, 2407–2419.
- Golubev, G.K., 2004. The method of risk envelope in estimation of linear functionals. *Problems Inform. Transmission* 40, 53–65.
- Golubev, G.K., Enikeeva, F., 2001. On the minimax estimation problem of a fractional derivative. *Theory Probab. Appl.* 46, 619–635.
- Golubev, G.K., Nussbaum, M., 1992. Adaptive spline estimates in a nonparametric regression model. *Theory Probab. Appl.* 37, 521–529.
- Kneip, A., 1993. Ordered linear smoothers. *Ann. Statist.* 21, 590–599.
- Mallows, A.L., 1973. Some comments on  $c_p$ . *Technometrics* 15, 661–675.
- Nussbaum, M., 1996. Asymptotic equivalence of density estimation and Gaussian white noise. *Ann. Statist.* 24, 2399–2430.
- Pinsker, M.S., 1980. Optimal filtering of square-integrable signals in Gaussian noise. *Problems Inform. Transmission* 16, 120–133.
- Stoyan, D., Kendall, W.S., Mecke, J., 1995. *Stochastic Geometry and its Applications*. Wiley, New York.
- Tsybakov, A.B., 2004. *Introduction à l'estimation non-paramétrique*. Springer, Berlin.
- Wicksell, S.D., 1925. The corpuscle problem. A mathematical study of a biometric problem. *Biometrika* 17, 84–99.
- Williams, D., 1991. *Probability with Martingales*. Cambridge University Press, Cambridge.
- Zygmund, A., 1968. *Trigonometric Series*. Cambridge University Press, Cambridge.