1. Introduction

In this paper, we consider a signal statistical treatment problem in the framework of a nonparametric regression model in continuous time, i.e.

\[ dy_t = S(t) dt + d\xi, \quad 0 \leq t \leq n \]  

where \( S(\cdot) \) is an unknown 1-periodic signal, \( \xi \) is an unobserved noise and \( n \) is the duration of observation. The problem is to estimate the function \( S \) on the observations \( y_t \). Note that if \( \xi \) is a Brownian motion, then we obtain the well-known "signal + white noise" model which is very popular in statistical radio-physics (see, for example, [1-3] and etc.). In this paper, we assume that in addition to the intrinsic noise in the radio-electronic system, approximated usually by the Gaussian white or color noise, the useful signal \( S \) is distorted by the impulse flow described by the Levy process, i.e. we assume that the noise process \( \xi \) is defined as

\[ \xi_t = \rho_1 w, + \rho_2 z_t \quad \text{and} \quad z_t = x_t (\mu - \mu) \]  

where \( \rho_1 \) and \( \rho_2 \) are some unknown constants, \( w \) is a standard Brownian motion, \( \mu(\cdot) \) is a jump measure with deterministic compensator \( \mu(\cdot) = \mu(dx) \), such that

\[ \Pi(x^+) = 1 \quad \text{and} \quad \Pi(x^-) < \infty \]  

Here we use the notation \( \Pi(|x|) = \int |y| \Pi(dy) \). Note that the Levy measure \( \Pi(\mathbb{R}) \) could be equal to \( +\infty \). We use \( * \) for the stochastic integrals with respect to random measures (see [4], Chs. 2 and 3), i.e.

\[ x * (\mu - \mu) = \int x(\mu - \mu)(ds, dy) \]  

In the sequel we will denote by \( Q \) the distribution of the process \( \xi \) and by \( Q_\gamma \) we denote all distributions for which the parameters \( \rho_1 \geq \gamma \) and \( \rho_1^2 + \rho_2^2 \leq \gamma^2 \) where \( \gamma \) and \( \gamma^2 \) are some fixed positive bounds. The cause of the appearance of a pulse stream in the radio-electronic systems can be, for example, either external unintended (atmospheric) or intentional impulse noise and the errors in the demodulation and the channel decoding for the binary information symbols. Note that, for the first time the impulse noises for detection signal problems have been introduced on the basis of compound Poisson processes was
by Kassam in [2]. However, the compound Poisson process can
describe only the large impulses influence of fixed single frequency.
There should be taking into account that in telecommunication
systems, impulses are without limitations on frequencies. So, one
needs to extend the framework of the observation model by
making use of the Levy processes as in Equations (2). In this
paper, we consider the estimation problem in the adaptive setting,
i.e. when the regularity of $S$ is unknown. Since the distribution $Q$
of the noise process $(\xi_j)_{j \in \mathbb{Z}}$ is unknown we use the robust
estimation approach developed for nonparametric problems in
[5]. We define the robust risk as

$$
\mathcal{R}^*(\hat{S}_n, S) = \sup_{Q \in \mathcal{D}} \mathcal{R}_Q(\hat{S}_n, S)
$$

where $\hat{S}_n$ is an estimation of the unknown signal $S$, i.e. some
function of $(y_i)_{i \in \mathbb{Z}}$, $\mathcal{R}_Q(\cdot, \cdot)$ is the usual quadratic risk defined as

$$
\mathcal{R}_Q(\hat{S}_n, S) = \mathbb{E}_Q \| \hat{S}_n - S \|^2 + \int_0^1 S'(t)dt
$$

Here $\mathbb{E}_Q$ stands for the expectation with respect to the
distribution $Q$ of the process in the Equation (1) with a fixed
distribution $Q$ of the noise $(\xi_j)_{j \in \mathbb{Z}}$ and a given function $S$.

In this paper, we consider minimax optimization criteria
which aim to minimize the robust risk which is defined in
the Equation (5) (see, for example, in [6]). To do this we use
the model selection methods. The interest to such statistical
procedures is explained by the fact that they provide adaptive
solutions for a nonparametric estimation through oracle
inequalities which give a non-asymptotic upper bound for a
quadratic risk including a minimal risk over chosen family of
estimators. It should be noted that the model selection methods
for parametric models were proposed, for the first time, by Akaike
[7]. Then, these methods had been developed by Barron, Birge
and Massart [8] and Fournierin and Pergamenshchikov [9]
for the nonparametric estimation and oracle inequalities for
the quadratic risks. Unfortunately, the oracle inequalities obtained
in these papers cannot provide the efficient estimation in the
adaptive setting, since the upper bounds in these inequalities
have some fixed coefficients in the main terms which are more
than one. In order to obtain the efficiency property for estimation
procedures, one has to obtain the sharp oracle inequalities, i.e.
in which the factor at the principal term on the right-hand side
of the inequality is close to unity. For this reason, one needs to
use the general semimartingale approach for the robust adaptive
efficient estimation of the nonparametric signals in continuous
time proposed by Konev and Pergamenshchikov [5]. The goal of
this paper is to develop a new sharp model selection method for
estimating the unknown signal $S$ using the improved estimation
approach. Usually, the model selection procedures are based on
the least square estimators. However, in this paper, we propose
to use the improved least square estimators which enable us to
considerably improve the non-asymptotic estimation accuracy.
Such idea was proposed, for the first time, in [9]. Our goal is to
develop these methods for non-Gaussian regression models in
continuous time and to obtain the sharp oracle inequalities. It
should be noted that to apply the improved estimation methods to
the non-Gaussian regression models in continuous time one needs
to modify the well-known James-Stein procedure introduced in
[10] in the way proposed in [11, 12]. So, by using these estimators
we construct the improved model selection procedure and we
show that the constructed estimation procedure is optimal in
the sense of the sharp non-asymptotic oracle inequalities for the
robust risks which are defined in the Equation (5).

2. Improved estimation

Let $(\phi_j)_{j \in \mathbb{Z}}$ be an orthonormal basis in $L_2[0, 1]$. We extend
these functions by the periodic way on $\mathbb{R}$, i.e. $\phi_j(t) = \phi_j(t+1)$
for any $t \in \mathbb{R}$. For estimating the unknown function $S$ in
the Equation (1) we consider it's Fourier expansion (see, for example,
[13])

$$
S(t) = \sum_{j \in \mathbb{Z}} \phi_j \theta_j(t) \quad \text{and} \quad \theta_j(t) = \frac{1}{n} \int_0^n S(t) \phi_j(t)dt
$$

The corresponding Fourier coefficients can be estimated as

$$
\hat{\theta}_j = \frac{1}{n} \int_0^n \phi_j(t)dy_i
$$

We define a class of weighted least squares estimates for $S(t)$ as

$$
\hat{S}_n = \sum_{j \in \mathbb{Z}} \lambda(j) \hat{\theta}_j \phi_j
$$

where the weights $\lambda \in \mathbb{R}^+$ belong to some finite set $\Lambda$ from
$[0, 1]^d$.

Now, for the first $d$ Fourier coefficients in Equations (7) we
use the improved estimation method proposed for parametric
models in [12]. To this end we set $\hat{\theta}_j = (\hat{\theta}_j)_{j \in \mathbb{Z}}$. In the sequel
we will use the norm $|x|^2 = \sum_{j \in \mathbb{Z}} x_j^2$ for any vector $x = (x_j)_{j \in \mathbb{Z}}$
from $\mathbb{R}^d$. Now we define the shrinkage estimators as

$$
\hat{\theta}_j = (1 - g(j)) \hat{\theta}_j
$$

where $g(j) = \frac{\alpha_j}{\alpha_j + 1}$, $\alpha_j$ is the indicator of the set $A$ and
$\alpha_j$ is some known parameter such that $\alpha_j \approx \frac{d}{n}$ as $n \to \infty$. Now
we introduce a class of shrinkage weighted least squares estimates for
$S$ as

$$
S_n = \sum_{j \in \mathbb{Z}} \lambda(j) \hat{\theta}_j \phi_j
$$
We denote the difference of quadratic risks of the estimates in Equation (11) and Equation (9) as $\Delta_n(S):=R_n(S')-R_n(S)$. Now for this deviation we obtain the following result.

**Theorem 1.** Assume that for any vector $\lambda \in \Lambda$ there exists some fixed integer $d = d(\Lambda)$ such that their first $d$ components equal to one, i.e. $\lambda(j) = 1$ for $1 \leq j \leq d$ for any $\lambda \in \Lambda$. Then for any $n \geq 1$ and $r>0$

$$\sup_{\lambda \in \Lambda} \Delta_n(S) < -c_i^2$$

(12)

The inequality (12) means that non-asymptotically, i.e. for any $n \geq 1$ the estimate in the Equation (11) outperforms in mean square accuracy the estimate in the Equation (9). Moreover, as we will see below, $n \rightarrow \infty$ as $\rightarrow \infty$. This means that improvement is considerable may be better than for the parametric regression [11].

3. Model selection

This Section gives the construction of a model selection procedure for estimating a function $S$ in the Equation (1) on the basis of improved weighted least square estimates and states the sharp oracle inequality for the robust risk of proposed procedure.

The model selection procedure for the unknown function $S$ in the Equation (1) will be constructed on the basis of a family of estimates $(S_i)_{i,\lambda}$.

The performance of any estimate $S_i$ will be measured by the empirical squared error

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

(13)

In order to obtain a good estimate, we have to write a rule to choose a weight vector $\lambda \in \Lambda$ in the Equation (6). It is obvious, that the best way is to minimize the empirical squared error with respect to $\lambda$. Making use the equation definition in the Equation (6) and the Fourier transformation of $S$ implies

$$\text{Err}_n(\lambda) = \sum_{j=1}^d \lambda(j) \theta_j^2 - 2 \sum_{j=1}^d \lambda(j) \theta_j \theta_j + \sum_{j=1}^d \theta_j^2$$

(14)

Since the Fourier coefficients $(\theta_j)_{j=1}^d$ are unknown, the weight coefficients $(\lambda_j)_{j=1}^d$ cannot be found by minimizing this quantity. To circumvent this difficulty one needs to replace the terms $\theta_j, \theta_j$ by their estimators $\hat{\theta}_j$. We set

$$\hat{\theta}_j = \theta_j - \frac{\delta_j}{n}$$

(15)

where $\delta_j$ is the estimate for the noise variance of $\sigma^2 = E_\sigma \xi_j^2 = \rho^2 + \rho^2$ which we choose in the following form

$$\delta_j = \sum_{j=1}^d \xi_j^2$$

and

$$\xi_j = \frac{1}{n} \int_{B} \Phi(t)dy_t$$

(16)

Here we denoted by $(\tilde{T}_r)_{r=1}^d$ the trigonometric basis in $L[0,1]$. For this change in the empirical squared error, one has to pay some penalty. Thus, one comes to the cost function of the form

$$J_n(\lambda) = \sum_{j=1}^d \lambda(j) \theta_j^2 - 2 \sum_{j=1}^d \lambda(j) \theta_j + \delta \lambda^2$$

(17)

where $\delta$ is some positive constant, $\delta \lambda$ is the penalty term defined as

$$\delta \lambda = \frac{\sigma^2}{n}$$

(18)

Substituting the weight coefficients, minimizing the cost function

$$\lambda^* = \arg \min_{\lambda \in \Lambda} J_n(\lambda)$$

(19)

in the Equation (11) leads to the improved model selection procedure

$$S^* = S_{\lambda^*}$$

(20)

It will be noted that $\lambda^*$ exists because $\Lambda$ is a finite set. If the minimizing sequence in the Equation (19) $\lambda^*$ is not unique, one can take any minimizer. In the case, when the value of $\sigma^2$ is known, one can take $\hat{\sigma} = \sigma^2$ and $\delta \lambda = \sigma^2 n^{-1}$.

**Theorem 2.** For any $n \geq 2$ and $0 < \delta < 1/3$, the robust risks defined in the Equation (5) of estimate in the Equation (20) for continuously differentiable function $S$ satisfies the oracle inequality

$$R(\lambda^*, S) \leq \frac{1 + 3\delta}{1 - 3\delta} \min_{\lambda \in \Lambda} R(S, S) + \frac{B}{n\delta}$$

(21)

where the rest term is such that $B_n^{-1} \rightarrow 0$ as $n \rightarrow \infty$ for any $\epsilon > 0$.

The inequality (21) means that the procedure in the Equation (20) is optimal in the oracle inequalities sense. This property enables to provide asymptotic efficiency in the adaptive setting, i.e. when information about the signal regularity is unknown.

4. Monte Carlo simulations

In this section we report the results of a Monte Carlo experiment to assess the performance of the proposed model selection procedure in the Equation (20). In the Equation (1) we choose 1-periodic function $S$ which, for $0 \leq t \leq 1$, is defined as $S(t) = \{t - \frac{1}{2}\}$, if $\frac{1}{4} \leq t \leq \frac{3}{4}$ and $S(t) = \frac{1}{2}$ elsewhere. We simulate the Equation (1) with the noise process defined as

$$\xi_t = 0.5\xi_t + 0.5e_t$$

(22)
since the accuracy improvement is proportional to the parameter dimension. We remember that for the nonparametric estimation this dimension tends to infinity, but in the parametric case it is always fixed. Therefore, the gain in the non-asymptotic quadratic accuracy from the application of the improved estimation methods is much more significant in statistical treatment problems of nonparametric signals. Moreover, as shown by the results of numerical simulation, the improved estimate for non-asymptotic accuracy exceeds the usual least squares estimators approximately in 1.5-2 times for the chosen experimental model.

Acknowledgement

The results of section 2 of this work are supported by the Ministry of Education and Science of the Russian Federation in the framework of the research project no. 2.3208.2017/4.6 and the RFBR Grant 16-01-00121 A. The results of Section 3 are supported by the RSF grant number 14-49-00079 (National Research University “MPEI”, Moscow, Russia). The last author is partially supported by the Russian Federal Professor program (project no. 1.472.2016/1.4, Ministry of Education and Science) and by the project XterM-Feder, University of Rouen.

Table 1 Empirical risks

<table>
<thead>
<tr>
<th>n</th>
<th>$\mathcal{R}(\hat{S},S)$</th>
<th>$\mathcal{R}(\hat{S}',S)$</th>
<th>$\mathcal{R}(\hat{S},S)/\mathcal{R}(\hat{S}',S)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>0.0462</td>
<td>0.0331</td>
<td>1.40</td>
</tr>
<tr>
<td>100</td>
<td>0.0262</td>
<td>0.0133</td>
<td>1.97</td>
</tr>
<tr>
<td>200</td>
<td>0.0126</td>
<td>0.00824</td>
<td>1.53</td>
</tr>
<tr>
<td>1000</td>
<td>0.00129</td>
<td>0.00094</td>
<td>1.37</td>
</tr>
</tbody>
</table>

Table 1 give the values for the sample risks for different numbers of observation period $n$.

5. Conclusion

In conclusion, we would like to emphasize that in this paper we developed new model selection procedures based on the improved versions of the least square estimators. It turns out that the improvement effect in the nonparametric estimation is more important than for the parameter estimation problems since the accuracy improvement is proportional to the parameter dimension. We remember that for the nonparametric estimation this dimension tends to infinity, but in the parametric case it is always fixed. Therefore, the gain in the non-asymptotic quadratic accuracy from the application of the improved estimation methods is much more significant in statistical treatment problems of nonparametric signals. Moreover, as shown by the results of numerical simulation, the improved estimate for non-asymptotic accuracy exceeds the usual least squares estimators approximately in 1.5-2 times for the chosen experimental model.

References


