



HAL
open science

Non asymptotic sharp oracle inequalities for the improved model selection procedures for the adaptive nonparametric signal estimation problem

Evgeny Pchelintsev, Valerii Pchelintsev, Serguei Pergamenchtchikov

► To cite this version:

Evgeny Pchelintsev, Valerii Pchelintsev, Serguei Pergamenchtchikov. Non asymptotic sharp oracle inequalities for the improved model selection procedures for the adaptive nonparametric signal estimation problem. *Komunikacie*, 2018. hal-02334888

HAL Id: hal-02334888

<https://hal.science/hal-02334888>

Submitted on 29 Oct 2019

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

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

Non asymptotic sharp oracle inequalities for the improved model selection procedures for the adaptive nonparametric signal estimation problem. *

Pchelintsev E.A., [†] Pchelintsev V.A., [‡]
Pergamenshchikov S.M. [§]

Abstract

In this paper, we consider the robust adaptive non parametric estimation problem for the periodic function observed with the Lévy noises in continuous time. An adaptive model selection procedure, based on the improved weighted least square estimates, is proposed. Sharp oracle inequalities for the robust risks have been obtained.

Key words: Improved non-asymptotic estimation, Weighted least squares estimates, Robust quadratic risk, Non-parametric regression, Lévy process, Model selection, Sharp oracle inequality, Asymptotic efficiency.

AMS (2010) Subject Classification : primary 62G08; secondary 62G05

*The section 2 of this work is 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 Section 3 is supported by the RSF grant number 14-49-00079 (National Research University "MPEI" 14 Krasnokazarmennaya, 111250 Moscow, Russia).

[†]Department of Mathematics and Mechanics, Tomsk State University, Lenin str. 36, 634050 Tomsk, Russia, e-mail: evgen-pch@yandex.ru

[‡]Department of Higher Mathematics and Mathematical Physics, Tomsk Polytechnic University, Lenin str. 30, 634050 Tomsk, Russia, e-mail: vpchelintsev@vtomske.ru

[§]Laboratoire de Mathématiques Raphael Salem, Avenue de l'Université, BP. 12, Université de Rouen, F76801, Saint Etienne du Rouvray, Cedex France and International Laboratory of Statistics of Stochastic Processes and Quantitative Finance of National Research Tomsk State University, e-mail: Serge.Pergamenchchikov@univ-rouen.fr

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.

$$dy_t = S(t)dt + d\xi_t, \quad 0 \leq t \leq n, \quad (1.1)$$

where $S(\cdot)$ is an unknown 1 - periodic signal, $(\xi_t)_{0 \leq t \leq n}$ is an unobserved noise and n is the duration of observation. The problem is to estimate the function S on the observations $(y_t)_{0 \leq t \leq n}$. Note that if $(\xi_t)_{0 \leq t \leq n}$ 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, [5, 8, 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 Lévy process, i.e. we assume that the noise process $(\xi_t)_{0 \leq t \leq n}$ is defined as

$$\xi_t = \varrho_1 w_t + \varrho_2 z_t \quad \text{and} \quad z_t = x * (\mu - \tilde{\mu})_t, \quad (1.2)$$

where, ϱ_1 and ϱ_2 are some unknown constants, $(w_t)_{t \geq 0}$ is a standard brownian motion, $\mu(ds dx)$ is a jump measure with deterministic compensator $\tilde{\mu}(ds dx) = ds\Pi(dx)$, $\Pi(\cdot)$ is a Lévy measure, i.e. some positive measure on $\mathbb{R}_* = \mathbb{R} \setminus \{0\}$, such that

$$\Pi(x^2) = 1 \quad \text{and} \quad \Pi(x^6) < \infty. \quad (1.3)$$

Here we use the notation $\Pi(|x|^m) = \int_{\mathbb{R}_*} |z|^m \Pi(dz)$. Note that the Lévy measure $\Pi(\mathbb{R}_*)$ could be equal to $+\infty$. In the sequel we will denote by Q the distribution of the process $(\xi_t)_{0 \leq t \leq n}$ and by \mathcal{Q}_n^* we denote all these distributions for which the parameters $\varrho_1 \geq \varsigma_*$ and $\varrho_1^2 + \varrho_2^2 \leq \varsigma^*$, where ς_* and ς^* 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 [8]. However, the compound Poisson process can describe only the large impulses influence of fixed single frequency. 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 Lévy processes (1.2). 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 $(\xi_t)_{0 \leq t \leq n}$ is unknown we use the robust estimation approach developed for nonparametric problems in [9]. To this end we define the robust risk as

$$\mathcal{R}^*(\widehat{S}_n, S) = \sup_{Q \in \mathcal{Q}_n^*} \mathcal{R}_Q(\widehat{S}_n, S) \quad (1.4)$$

where \widehat{S}_n is an estimation, i.e. any function of $(y_t)_{0 \leq t \leq n}$, $\mathcal{R}_Q(\cdot, \cdot)$ is the usual quadratic risk defined as

$$\mathcal{R}_Q(\widehat{S}_n, S) := \mathbf{E}_{Q,S} \|\widehat{S}_n - S\|^2 \quad \text{and} \quad \|S\|^2 = \int_0^1 S^2(t) dt. \quad (1.5)$$

In this paper, we consider a minimax optimisation criteria which aims to minimize the robust risk (1.4) (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 [1]. Then, these methods had been developed by Barron, Birgé and Massart [2] and Fourdrinier and Pergamenshchikov [4] for the nonparametric estimation and oracle inequalities for the quadratic risks. Unfortunately, the oracle inequalities obtained in these papers can not 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 semi - martingale approach for the robust adaptive efficient estimation of the nonparametric signals in continuous time proposed by Konev and Pergamenshchikov in [9]. 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 [4]. 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 [7] in the way proposed in [11, 10]. 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 (1.4).

2 Improved estimation

Let $(\phi_j)_{j \geq 1}$ be an orthonormal basis in $\mathbf{L}_2[0, 1]$. 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 (1.1) we consider it's Fourier expansion

$$S(t) = \sum_{j=1}^{\infty} \theta_j \phi_j(t) \quad \text{and} \quad \theta_j = (S, \phi_j) = \int_0^1 S(t) \phi_j(t) dt. \quad (2.1)$$

The corresponding Fourier coefficients can be estimated as

$$\widehat{\theta}_{j,n} = \frac{1}{n} \int_0^n \phi_j(t) dy_t. \quad (2.2)$$

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

$$\widehat{S}_\lambda = \sum_{j=1}^n \lambda(j) \widehat{\theta}_{j,n} \phi_j, \quad (2.3)$$

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

Now, for the first d Fourier coefficients in (2.1) we use the improved estimation method proposed for parametric models in [11]. To this end we set $\widetilde{\theta}_n = (\widehat{\theta}_{j,n})_{1 \leq j \leq d}$. In the sequel we will use the norm $|x|_d^2 = \sum_{j=1}^d x_j^2$ for any vector $x = (x_j)_{1 \leq j \leq d}$ from \mathbb{R}^d . Now we define the shrinkage estimators as

$$\theta_{j,n}^* = (1 - g(j)) \widehat{\theta}_{j,n}, \quad (2.4)$$

where $g(j) = (\mathbf{c}_n / |\widetilde{\theta}_n|_d) \mathbf{1}_{\{1 \leq j \leq d\}}$, and \mathbf{c}_n is some known parameter such that $\mathbf{c}_n \approx d/n$ as $n \rightarrow \infty$. Now we introduce a class of shrinkage weighted least squares estimates for S as

$$S_\lambda^* = \sum_{j=1}^n \lambda(j) \theta_{j,n}^* \phi_j. \quad (2.5)$$

We denote the difference of quadratic risks of the estimates (2.3) and (2.5) as $\Delta_Q(S) := \mathcal{R}_Q(S_\lambda^*, S) - \mathcal{R}_Q(\widehat{S}_\lambda, S)$. Now for this deviation we obtain the following result.

Theorem 2.1. *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 $n \geq 1$*

$$\sup_{Q \in \mathcal{Q}_n} \sup_{\|S\| \leq r} \Delta_Q(S) < -\mathbf{c}_n^2. \quad (2.6)$$

The inequality (2.6) means that non asymptotically, i.e. for any $n \geq 1$ the estimate (2.5) outperforms in mean square accuracy the estimate (2.3). Moreover, as we will see below, $n\mathbf{c}_n \rightarrow \infty$ as $d \rightarrow \infty$. This means that improvement is considerable may better than for the parametric regression (see, [11]).

3 Model selection

This Section gives the construction of a model selection procedure for estimating a function S in (1.1) 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 (1.1) will be constructed on the basis of a family of estimates $(S_\lambda^*)_{\lambda \in \Lambda}$.

The performance of any estimate S_λ^* will be measured by the empirical squared error

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

In order to obtain a good estimate, we have to write a rule to choose a weight vector $\lambda \in \Lambda$ in (2.5). It is obvious, that the best way is to minimise the empirical squared error with respect to λ . Making use the estimate definition (2.5) and the Fourier transformation of S implies

$$\text{Err}_n(\lambda) = \sum_{j=1}^n \lambda^2(j) (\theta_{j,n}^*)^2 - 2 \sum_{j=1}^n \lambda(j) \theta_{j,n}^* \theta_j + \sum_{j=1}^n \theta_j^2. \quad (3.1)$$

Since the Fourier coefficients $(\theta_j)_{j \geq 1}$ are unknown, the weight coefficients $(\lambda_j)_{j \geq 1}$ can not be found by minimizing this quantity. To circumvent this difficulty one needs to replace the terms $\theta_{j,n}^* \theta_j$ by their estimators $\tilde{\theta}_{j,n}$. We set

$$\tilde{\theta}_{j,n} = \theta_{j,n}^* \hat{\theta}_{j,n} - \frac{\hat{\sigma}_n}{n}, \quad (3.2)$$

where $\widehat{\sigma}_n$ is the estimate for the noise variance of $\sigma_Q = \mathbf{E}_Q \xi_{j,n}^2 = \varrho_1^2 + \varrho_2^2$ which we choose in the following form

$$\widehat{\sigma}_n = \sum_{j=\lfloor \sqrt{n} \rfloor + 1}^n \widehat{t}_{j,n}^2 \quad \text{and} \quad \widehat{t}_{j,n} = \frac{1}{n} \int_0^n \text{Tr}_j(t) dy_t.$$

Here we denoted by $(\text{Tr}_j)_{j \geq 1}$ the trigonometric basis in $\mathbf{L}_2[0, 1]$. For this change in the empirical squared error, one has to pay some penalty. Thus, one comes to the cost function of the form

$$J_n(\lambda) = \sum_{j=1}^n \lambda^2(j) (\theta_{j,n}^*)^2 - 2 \sum_{j=1}^n \lambda(j) \widetilde{\theta}_{j,n} + \delta \widehat{P}_n(\lambda), \quad (3.3)$$

where δ is some positive constant, $\widehat{P}_n(\lambda)$ is the penalty term defined as

$$\widehat{P}_n(\lambda) = \frac{\widehat{\sigma}_n |\lambda|_n^2}{n}. \quad (3.4)$$

Substituting the weight coefficients, minimizing the cost function

$$\lambda^* = \operatorname{argmin}_{\lambda \in \Lambda} J_n(\lambda), \quad (3.5)$$

in (2.3) leads to the improved model selection procedure

$$S^* = S_{\lambda^*}^*. \quad (3.6)$$

It will be noted that λ^* exists because Λ is a finite set. If the minimizing sequence in (3.5) λ^* is not unique, one can take any minimizer. In the case, when the value of σ_Q is known, one can take $\widehat{\sigma}_n = \sigma_Q$ and $P_n(\lambda) = \sigma_Q |\lambda|_n^2/n$.

Theorem 3.1. *For any $n \geq 2$ and $0 < \delta < 1/2$, the robust risks (1.4) of estimate (3.6) for continuously differentiable function S satisfies the oracle inequality*

$$\mathcal{R}^*(S_{\lambda^*}^*, S) \leq \frac{1 + 2\delta}{1 - 2\delta} \min_{\lambda \in \Lambda} \mathcal{R}^*(S_\lambda^*, S) + \frac{B_n^*}{n\delta}, \quad (3.7)$$

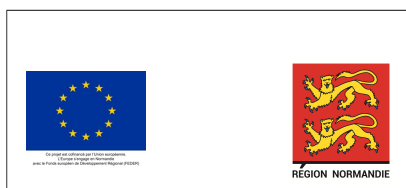
where the rest term is such that $B_n^*/n^\epsilon \rightarrow 0$ as $n \rightarrow \infty$ for any $\epsilon > 0$.

The inequality (3.7) means that the procedure (3.6) 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 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.

Acknowledgements. 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.



References

- [1] Akaike H. A new look at the statistical model identification. *IEEE Trans. on Automatic Control* **19** (1974) 716–723.
- [2] Barron, A. , Birgé, L. and Massart, P. (1999) Risk bounds for model selection via penalization. *Probab. Theory Relat. Fields* **113**, 301–415.
- [3] Chernoyarov O.V., Vaculik M., Shirikyan A., Salnikova A.V. (2015) Statistical analysis of fast fluctuating random signals with arbitrary function envelope and unknown parameters. *Komunikacie*, 17 (1A), 35–43.
- [4] Fourdrinier, D. and Pergamenshchikov, S. (2007) Improved selection model method for the regression with dependent noise. *Annals of the Institute of Statistical Mathematics*, **59**(3), 435–464.
- [5] Ibragimov I. A. and Khasminskii R. Z. *Statistical Estimation: Asymptotic Theory*. Springer, New York, 1981.

- [6] Janacek J., Kvet M. (2015) Min-max optimization of emergency service system by exposing constraints. *Komunikacie*, 17 (2), 15–22.
- [7] James, W., Stein, C. (1961). Estimation with quadratic loss. In *Proceedings of the Fourth Berkeley Symposium Mathematics, Statistics and Probability*, University of California Press, Berkeley, 1, 361–380
- [8] Kassam, S.A. Signal detection in non-Gaussian noise. – New York: Springer-Verlag Inc., IX, 1988.
- [9] Konev V. V. and Pergamenshchikov S. M. Robust model selection for a semimartingale continuous time regression from discrete data. *Stochastic processes and their applications*, **125**, 2015, 294 – 326.
- [10] Konev V., Pergamenshchikov, S. and Pchelintsev, E. (2014) Estimation of a regression with the pulse type noise from discrete data. *Theory Probab. Appl.*, **58** (3), 442–457.
- [11] Pchelintsev, E. (2013) Improved estimation in a non-Gaussian parametric regression. *Stat. Inference Stoch. Process.*, **16** (1), 15 – 28.