DOI 10.1007/s10958-019-04209-w
Journal of Mathematical Sciences, Vol. 237, No. 6, March, 2019

AVERAGED PROBABILITY OF THE ERROR IN CALCULATING WAVELET COEFFICIENTS FOR THE RANDOM SAMPLE SIZE

O. V. Shestakov[1]

Signal denoising methods based on the threshold processing of wavelet coefficients are widely used in various application areas. When applying these methods, it is usually assumed that the number of wavelet coefficients is fixed, and the noise distribution is Gaussian. Such a model has been well studied in the literature, and optimal threshold values have been calculated for different signal classes and loss functions. However, in some situations the sample size is not known in advance and is modeled by a random variable. In this paper, we consider a model with a random number of observations contaminated by a Gaussian noise, and study the behavior of the loss function based on the probabilities of errors in calculating wavelet coefficients for a growing sample size.

1. Introduction

In many areas, such as plasma physics, medicine, geophysics, astronomy, and communication systems there is a need for analyzing and processing signals of a different kind and origin. One of the first steps of data processing is a transformation leading to their "economical" representation. One possible example of such transformation is a wavelet transform that exploits the correlations between adjacent samples in a digital signal, to a sparse data representation. This principle is also the basis for a popular threshold methods to reduce noise in signals: small wavelet coefficients are assumed to be dominated by noise and carry little useful information. Replacing these coefficients by zero eliminates a major part of the noise without affecting the signal too much. These procedures also compress the data with little loss of information, which allows one to store information more economically and transfer it faster through digital communication channels.

In some cases, the amount of data available for analysis (sample size) is not known in advance. Such situations can arise, for example, in the case of missing data, limited data acquisition time at random recording times, or lack of information on the characteristics of the equipment used. For example, the device used may belong to a batch within which certain specifications may be not rigidly fixed. In such a case, it is assumed that the sample size of the data is a random variable with some probability distribution.

In models with a fixed sample size, the statistical properties of wavelet threshold methods are well studied, and expressions for the "optimal" thresholds oriented to various loss functions are obtained [1–6]. In this paper, we consider a model with a random number of wavelet coefficients of a signal function "contaminated" with a white Gaussian noise. A loss function based on the probabilities of errors in the calculation of wavelet coefficients is considered, and its order is estimated.

2. Data model

Let the function of the observed signal belong to a class of functions possessing a certain degree of smoothness. After the wavelet transform of the observed signal, a set of empirical wavelet coefficients is obtained. In this paper we assume that these coefficients have the following form:

$$Y_{j,k} = \mu_{j,k} + z_{j,k}, \quad j = 0, \dots, J-1, \quad k = 0, \dots, 2^j - 1, \tag{1}$$

[1] Department of Mathematical Statistics, Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University, Moscow, Russia; Institute of Informatics Problems, Federal Research Center "Computer Science and Control" of the Russian Academy of Sciences, Moscow, Russia, e-mail: oshestakov@cs.msu.su

Proceedings of the XXXIV International Seminar on Stability Problems for Stochastic Models, Debrecen, Hungary.

where $\mu_{j,k}$ are the wavelet coefficients of the "pure" signal and $z_{j,k}$ are the "noise" wavelet coefficients for which it is assumed that they are independent and have a normal distribution with zero mean and variance σ^2. Since the signal function has a certain degree of smoothness, the absolute values of its wavelet coefficients decay with increasing j. Suppose that for some $\gamma > 0$ there exists a positive constant A such that

$$|\mu_{j,k}| \leqslant \frac{A2^{J/2}}{2^{j(\gamma+1/2)}}. \tag{2}$$

Such a rate of decay is provided, for example, by the membership of the signal function in the class of Lipschitz-regular functions with regularity exponent γ, provided that the basic wavelet functions also satisfy certain additional smoothness conditions [7]. We denote by $C(A,\gamma)$ the class of functions whose wavelet coefficients satisfy (2).

A popular denoising method is the threshold processing of empirical wavelet coefficients, the meaning of which is to remove the coefficients whose absolute values do not exceed a given threshold. The estimate $\hat{Y}_{j,k}$ is calculated using the threshold function $\rho_T(Y_{j,k})$ with the threshold T. The most common function is the hard thresholding function $\rho_T^{(h)}(x) = x\mathbf{1}(|x| > T)$ and soft thresholding function $\rho_T^{(s)}(x) = \operatorname{sign}(x)(|x| - T)_+$. The meaning of this processing is that since most of the "pure" coefficients are small in absolute value, the zeroing of the empirical coefficients should remove the noise without greatly affecting the useful signal.

3. The loss function for a nonrandom number of wavelet coefficients

A model in which the number of empirical wavelet coefficients is not random has been well studied. For it optimal thresholds are calculated and the orders of various loss functions are estimated [1–6]. In particular, the function [5] considers the loss function

$$r_J(f) = \mathsf{EP}\left(\left|\hat{Y}_{\xi,\eta} - \mu_{\xi,\eta}\right| > \varepsilon \mid \xi,\eta\right) = \frac{\sum_{j=0}^{J-1}\sum_{k=0}^{2^j-1} \mathsf{P}\left(\left|\hat{Y}_{j,k} - \mu_{j,k}\right| > \varepsilon\right)}{2^J}. \tag{3}$$

It represents the average probability that the error in computing the wavelet coefficient will exceed ε. This definition of the loss function is a generalization of the definition proposed in [4]. In the same paper [4] it was shown that the estimates whose purpose is to minimize losses $r_J(f)$ give comparable, and sometimes better, results than estimates minimizing the mean-square risk.

In [5] the following statements that estimate the minimax order of (3) are proved.

Theorem 1. *When choosing the asymptotically optimal threshold for a hard threshold processing the following inequalities are valid:*

$$C_m 2^{-\frac{2\gamma}{2\gamma+1}J} J^{-\frac{1}{2\gamma+1}} \leqslant \sup_{f\in C(A,\gamma)} r_J(f) \leqslant C_M 2^{-\frac{2\gamma}{2\gamma+1}J} J^{\frac{1}{2\gamma+1}}.$$

Theorem 2. *When choosing the asymptotically optimal threshold for a soft threshold processing the following inequalities are valid:*

$$C_m 2^{-\frac{2\gamma}{2\gamma+1}J} g_1(J) \leqslant \sup_{f\in C(A,\gamma)} r_J(f) \leqslant C_M 2^{-\frac{2\gamma}{2\gamma+1}J} g_2(J).$$

Here C_m and C_M are some positive constants, the function $g_1(J) > 0$ tends to zero arbitrarily slowly, and $g_2(J) > 0$ increases unboundedly in J such that

$$\ln g_2(J) = o(\sqrt{\ln 2^J}), \qquad J \to \infty.$$

The asymptotically optimal threshold in Theorems 1 and 2 for $J \to \infty$ satisfies the relation [5]

$$T \simeq \sigma\sqrt{\frac{4\gamma}{2\gamma+1}\ln 2^J}.$$

In the next section, the order of the loss function in the model with a random number of empirical wavelet coefficients is estimated.

4. The random number of wavelet coefficients

We now consider a model with a random number of wavelet coefficients. Let M be a positive integer random variable and $N = 2^M$. Then the analogue of the loss function (3) for the model (1) takes the form

$$r(f) = \mathsf{EP}\left(\left|\hat{Y}_{\xi,\eta} - \mu_{\xi,\eta}\right| > \varepsilon \mid \xi, \eta, N\right) = \sum_{J=0}^{\infty} \mathsf{P}\left(N = 2^J\right) \frac{\sum_{j=0}^{J-1}\sum_{k=0}^{2^j-1} \mathsf{P}\left(\left|\hat{Y}_{j,k} - \mu_{j,k}\right| > \varepsilon\right)}{2^J} \tag{4}$$

and its asymptotic order depends to a large extent on the distribution of N. To obtain meaningful estimates of the order of the loss function (4), N must be "large." Consider the sequence N_n, $n = 1, \ldots$, and suppose that there exists a nonrandom increasing sequence of natural numbers J_n, $n = 1, \ldots$, such that $N_n/2^{J_n}$ has a certain limit (in the sense of uniform convergence in the distribution) when $n \to \infty$, that is,

$$\sup_{x \geqslant 0} |H_n(x) - H(x)| < \frac{\varepsilon_n}{2} \to 0, \qquad n \to \infty, \tag{5}$$

where

$$H_n(x) = \mathsf{P}\left(\frac{N_n}{2^{J_n}} < x\right),$$

and $H(x)$ is the limit distribution function. Suppose that $H(x)$ has no atom at zero and let us investigate the behavior

$$r_n(f) = \sum_{J=0}^{\infty} \mathsf{P}\left(N_n = 2^J\right) \frac{\sum_{j=0}^{J-1}\sum_{k=0}^{2^j-1} \mathsf{P}\left(\left|\hat{Y}_{j,k} - \mu_{j,k}\right| > \varepsilon\right)}{2^J}$$

when $n \to \infty$.

Let $\delta_n \to 0$, $\alpha_n \to 0$ when $n \to \infty$ such that $J_n + \log_2 \delta_n \to \infty$ and $H(\delta_n) + 1 - H(\delta_n^{-1}) < \alpha_n$ for all $n = 1, \ldots$. Then

$$\begin{aligned} r_n(f) = &\sum_{J=0}^{[J_n+\log_2 \delta_n]} \mathsf{P}\left(N_n = 2^J\right) \frac{\sum_{j=0}^{J-1}\sum_{k=0}^{2^j-1} \mathsf{P}\left(\left|\hat{Y}_{j,k} - \mu_{j,k}\right| > \varepsilon\right)}{2^J} + \\ &+ \sum_{J=[J_n+\log_2 \delta_n]+1}^{[J_n-\log_2 \delta_n]} \mathsf{P}\left(N_n = 2^J\right) \frac{\sum_{j=0}^{J-1}\sum_{k=0}^{2^j-1} \mathsf{P}\left(\left|\hat{Y}_{j,k} - \mu_{j,k}\right| > \varepsilon\right)}{2^J} + \\ &+ \sum_{J=[J_n-\log_2 \delta_n]+1}^{\infty} \mathsf{P}\left(N_n = 2^J\right) \frac{\sum_{j=0}^{J-1}\sum_{k=0}^{2^j-1} \mathsf{P}\left(\left|\hat{Y}_{j,k} - \mu_{j,k}\right| > \varepsilon\right)}{2^J} \equiv S_1 + S_2 + S_3. \end{aligned}$$

Given (5), for $S_1 + S_3$ we have

$$S_1 + S_3 \leqslant H_n(\delta_n) + 1 - H_n(\delta_n^{-1}) \leqslant \alpha_n + \varepsilon_n.$$

When using a hard threshold processing and choosing the asymptotically optimal threshold, for S_2 with the use of Theorem 1, we obtain the estimate

$$S_2 \leqslant C_1 2^{-\frac{2\gamma}{2\gamma+1}(J_n+\log_2\delta_n)}(J_n+\log_2\delta_n)^{\frac{1}{2\gamma+1}},$$

valid for all $f \in C(A,\gamma)$, and for a soft threshold processing with the use of Theorem 2 we obtain the estimate

$$S_2 \leqslant C_2 2^{-\frac{2\gamma}{2\gamma+1}(J_n+\log_2\delta_n)} g_2(J_n+\log_2\delta_n),$$

where C_1 and C_2 are some positive constants. Thus, the following statement holds.

Theorem 3. *In the model with a random number of wavelet coefficients, when using a hard threshold processing and choosing the asymptotically optimal threshold, starting with some n we have the following estimate*:

$$\sup_{f\in C(A,\gamma)} r_n(f) \leqslant \alpha_n + \varepsilon_n + C_1 2^{-\frac{2\gamma}{2\gamma+1}(J_n+\log_2\delta_n)}(J_n+\log_2\delta_n)^{\frac{1}{2\gamma+1}},$$

and for a soft threshold processing

$$\sup_{f\in C(A,\gamma)} r_n(f) \leqslant \alpha_n + \varepsilon_n + C_2 2^{-\frac{2\gamma}{2\gamma+1}(J_n+\log_2\delta_n)} g_2(J_n+\log_2\delta_n).$$

The asymptotically optimal threshold itself for $n\to\infty$ satisfies the relation

$$T_n \simeq \sigma\sqrt{\frac{4\gamma}{2\gamma+1}\ln 2^{J_n+\log_2\delta_n}}.$$

The form of α_n, ε_n and δ_n in Theorem 3 depends essentially on the behavior of the sequence $N_n/2^{J_n}$ and the limit distribution function $H(x)$. Thus ε_n characterizes the convergence rate of $H_n(x)$ to $H(x)$, and α_n, δ_n depend on the behavior of $H(x)$ in a neighborhood of zero and infinity.

Corollary 1. *If the limit distribution of $N_n/2^{J_n}$ is degenerate*: *$N_n/2^{J_n} \xrightarrow{\mathsf{P}} 1$ when $n\to\infty$, then starting with some n*

$$\sup_{f\in C(A,\gamma)} r_n(f) \leqslant \varepsilon_n + C_3 2^{-\frac{2\gamma}{2\gamma+1}J_n} J_n^{\frac{1}{2\gamma+1}}$$

for a hard threshold processing, *and*

$$\sup_{f\in C(A,\gamma)} r_n(f) \leqslant \varepsilon_n + C_4 2^{-\frac{2\gamma}{2\gamma+1}J_n} g_2(J_n)$$

for a soft threshold processing, *where C_3 and C_4 are some positive constants.*

If ε_n decays rapidly enough, then these estimates coincide with the estimates for the loss function with a nonrandom number of wavelet coefficients and, in addition, lower minimax estimates similar to those given in Theorems 1 and 2 are valid.

Corollary 2. *Let $H(x)$ be differentiable in a neighborhood of zero and in this neighborhood $b \leqslant H'(x) \leqslant B$ for some positive constants b and B. Let $\delta_n = 2^{-\frac{2\gamma}{4\gamma+1}J_n}$. Then, starting with some n,*

$$\sup_{f\in C(A,\gamma)} r_n(f) \leqslant \varepsilon_n + C_5 2^{-\frac{2\gamma}{4\gamma+1}J_n} J_n^{\frac{1}{2\gamma+1}} \tag{6}$$

for a hard threshold processing, *and*

$$\sup_{f\in C(A,\gamma)} r_n(f) \leqslant \varepsilon_n + C_6 2^{-\frac{2\gamma}{4\gamma+1}J_n} g_2(J_n) \tag{7}$$

for a soft threshold processing, *where C_5 and C_6 are some positive constants.*

Note that $\frac{2\gamma}{4\gamma+1} < \frac{1}{2}$ for $\gamma > 0$, and if $\varepsilon_n = O(2^{-J_n/2})$ (which is a common estimate for the rate of convergence in the distribution), then the second terms in (6) and (7) determine the rate of decay of the loss function. Thus, the loss function for a random number of wavelet coefficients can tend to zero much slower than the loss function for a nonrandom number of wavelet coefficients.

Acknowledgements

This research is supported by Russian Science Foundation, project No. 18–11–00155.

REFERENCES

1. D. Donoho and I. M. Johnstone, "Ideal spatial adaptation via wavelet shrinkage," *Biometrika*, **81**, No. 3, 425–455 (1994).
2. D. Donoho and I. M. Johnstone, "Minimax estimation via wavelet shrinkage," *Ann. Stat.*, **26**, No. 3, 879–921 (1998).
3. M. Jansen, *Noise Reduction by Wavelet Thresholding*, Springer, New York (2001).
4. J. Sadasiva, S. Mukherjee, and C. S. Seelamantula, "An optimum shrinkage estimator based on minimum-probability-of-error criterion and application to signal denoising," in: *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Piscataway, New Jersey (2014), pp. 4249–4253.
5. A. A. Kudryavtsev and O. V. Shestakov, "Asymptotic behavior of the threshold minimizing the average probability of error in calculation of wavelet coefficients," *Dokl. Math.*, **93**, No. 3, 295–299 (2016).
6. A. A. Kudryavtsev and O. V. Shestakov, "Asymptotically optimal wavelet thresholding in the models with non-Gaussian noise distributions," *Dokl. Math.*, **94**, No. 3, 615–619 (2016).
7. S. Mallat, *A Wavelet Tour of Signal Processing*, Academic Press, New York (1999).