DOI 10.1007/s10958-020-04756-7
Journal of Mathematical Sciences, Vol. 246, No. 4, April, 2020

ON THE CHOICE OF THRESHOLDING PARAMETERS FOR NON-GAUSSIAN NOISE DISTRIBUTION

A. A. Kudryavtsev[1] **and O. V. Shestakov**[1,2]

The paper considers the problem of estimating the signal function from noisy observations using threshold processing of its wavelet expansion coefficients. Under general assumptions about the properties of the noise distribution, the asymptotic order of the optimal threshold is calculated, minimizing the loss function, based on the probability that the maximum error in the wavelet coefficients exceeds a given critical level.

1. Introduction

Wavelet methods are widely used in signal analysis and processing. Their main advantages are the speed of algorithms and the ability to locally process the signal functions in both the time (space) and frequency domain. It is usually assumed that the wavelet decomposition of the processed function contains only a small number of wavelet coefficients, which are large in their absolute value. In the problem of noise suppression, this assumption serves as a justification for the application of thresholding methods in which all wavelet coefficients, the absolute value of which is less than a certain threshold, are set to zero. The case where the noise in the signal is additive and Gaussian is well studied and expressions are obtained for optimal or asymptotically optimal thresholds oriented to different loss functions [1–6]. In this paper, it is assumed that the signal function belongs to the Lipschitz class with a certain positive exponent, and only the most general assumptions are made regarding the noise distribution. Relations are obtained that allow to calculate the threshold, which provides the asymptotically optimal order of the loss function, based on the probability that the maximum error in the wavelet coefficients exceeds a given critical level. Examples of the calculation of the asymptotically optimal threshold for various noise distributions are given.

2. Statement of the problem of finding an asymptotically optimal threshold

Let the signal function f be given on a finite interval $[a,b]$ and $f \in \mathrm{Lip}(\gamma, L)$, where $\mathrm{Lip}(\gamma, L)$ is the class of uniformly Lipschitz-regular functions on $[a,b]$ with some exponent $\gamma > 0$ and Lipschitz constant $L > 0$. In practice, f is given in N discrete samples. In this paper, it is assumed that $N = 2^J$ for some $J > 0$. After applying a discrete wavelet transform, the form of which is determined by a certain wavelet function ψ, we get a set of wavelet coefficients $\{\mu_{j,k}\}_{j=0,\ldots,J-1,\, k=0,\ldots,2^j-1}$, in which the index j is called the scale, and the index k is called the shift [7].

We impose some additional conditions on the wavelet function ψ: let ψ have M vanishing moments ($M \geqslant \gamma$), be M times continuously differentiable, and for all $0 \leqslant k \leqslant M$ and any $m \in \mathbf{N}$ let there be a constant C_m such that for all $x \in \mathbf{R}$

$$\left|\psi^{(k)}(x)\right| \leqslant \frac{C_m}{1+|x|^m}.$$

[1] Department of Mathematical Statistics, Faculty of Computational Mathematics and Cybernetics, M. V. Lomonosov Moscow State University, e-mail: `nubigena@mail.ru`

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

Proceedings of the XXXV International Seminar on Stability Problems for Stochastic Models, Perm, Russia, September 24–28, 2018. Part I.

Then there is a constant $C_f > 0$ such that [7]

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

In real observations, there is always a noise due to imperfections of equipment, various interference and other reasons. For empirical wavelet coefficients in this paper, the following noise signal model is adopted:

$$Y_{j,k} = \mu_{j,k} + W_{j,k}, \quad j = 0, \ldots, J-1, \; k = 0, \ldots, 2^j - 1,$$

where $\mu_{j,k}$ are discrete wavelet coefficients of the "pure" signal f, and $W_{j,k}$ are the "noise" coefficients which are assumed to be independent and have a symmetric absolutely continuous differentiable distribution function $\mathsf{P}(W_{j,k} < x) = 1 - H(x)$. We suppose that $0 < H(x) < 1$ for all $x \in \mathbf{R}$.

When applying the wavelet methods for noise suppression, the threshold processing method is most frequently used. It sets to zero the coefficients whose absolute values do not exceed a given threshold T, since it is believed that the main part of the useful signal is contained in a relatively small number of relatively large coefficients. The threshold processing function ρ_T is applied to each coefficient. This paper considers most commonly used functions of hard thresholding $\rho_T^{(h)}(x) = x \cdot \mathbf{1}(|x| > T)$ and soft thresholding $\rho_T^{(s)}(x) = \mathrm{sign}(x)(|x| - T)_+$. We denote by $\hat{Y}_{j,k}$ the estimate of the wavelet coefficient, which is obtained using the threshold processing: $\hat{Y}_{j,k} = \rho_T(Y_{j,k})$.

Let some critical value $\varepsilon > 0$ be given. Consider the loss function of the following form:

$$l_J(f) = \mathsf{P}\left(\max_{j,k} \left|\hat{Y}_{j,k} - \mu_{j,k}\right| > \varepsilon\right). \tag{2}$$

This function is a generalization of the loss function proposed in [4]. In the same paper [4], it was shown that the choice of thresholds, the purpose of which is to minimize the loss function based on the error probabilities, gives comparable and sometimes better results than thresholds that minimize the mean-square error.

With an unbounded increase of the number of samples, $l_J(f)$ tends to unity. The asymptotically optimal threshold should provide minimal losses in the sense that the speed of $l_J(f)$ tending to unity for the given threshold is the smallest. Due to the relation

$$1 - \mathsf{P}\left(\max_{j,k} \left|\hat{Y}_{j,k} - \mu_{j,k}\right| > \varepsilon\right) = \prod_{j=0}^{J-1} \prod_{k=0}^{2^j-1} \mathsf{P}\left(\left|\hat{Y}_{j,k} - \mu_{j,k}\right| \leqslant \varepsilon\right)$$

it is equivalent to the fact that for the asymptotically optimal threshold, the speed of convergence to infinity of the loss function of the form

$$r_J(f) = \sum_{j=0}^{J-1} \sum_{k=0}^{2^j-1} \left|\ln \mathsf{P}\left(\left|\hat{Y}_{j,k} - \mu_{j,k}\right| \leqslant \varepsilon\right)\right| \tag{3}$$

is the slowest.

The goal of this paper is to derive relations that allow us to calculate the asymptotically optimal threshold in the class $\mathrm{Lip}(\gamma, L)$ for the loss function, defined as

$$R_J = \sup_{f \in \mathrm{Lip}(\gamma, L)} r_J(f), \tag{4}$$

i.e., the threshold asymptotically optimal in the minimax sense. A detailed study of the behavior of the asymptotically optimal threshold for the mean-square loss function in the model with additive Gaussian noise can be found in [2] and [3]. Also in [1], a method was proposed for finding an adaptive optimal threshold with which one can estimate the mean-square error of the threshold processing of a specific function. This method is based on minimizing the unbiased risk estimate, the statistical properties of which are studied in detail in the papers [8] and [9]. Note that the "reasonable" threshold should increase in J [3]. However, in order to simplify the notation, the dependence of the threshold on J will not be explicitly indicated below.

In what follows, the symbol $\asymp$ denotes the order of the quantity in question in J, i.e., $a_J \asymp b_J$ if, starting with some J, the inequalities $C_1 \cdot b_J \leqslant a_J \leqslant C_2 \cdot b_J$ hold for some positive constants C_1 and C_2. The notation $a_J \sim b_J$ will be used if $\lim\limits_{J\to\infty} a_J/b_J = 1$.

3. Hard thresholding

Let the estimates of the wavelet coefficients be calculated using hard threshold processing: $\hat{Y}_{j,k} = \rho_T^{(h)}(Y_{j,k})$.

Consider the loss function (4). Note that for an arbitrary $\varepsilon > 0$ there is a function $f \in \mathrm{Lip}(\gamma, L)$, such that the inequality (1) turns to the equality for some j_h such that $|\mu_{j_h,k}| > \varepsilon$. Therefore, there is $J_0 > 0$ such that for $J > J_0$, $|\mu_{j_h,k}| > \varepsilon$, $T - \mu_{j_h,k} > \varepsilon$ and $-T - \mu_{j_h,k} < -\varepsilon$. In this case

$$\mathsf{P}\left(\left|\hat{Y}_{j_h,k} - \mu_{j_h,k}\right| \leqslant \varepsilon\right) = \mathsf{P}(-\varepsilon \leqslant Y_{j_h,k} - \mu_{j_h,k} \leqslant \varepsilon, Y_{j_h,k} - \mu_{j_h,k} > T - \mu_{j_h,k}) +$$
$$+ \mathsf{P}(-\varepsilon \leqslant Y_{j_h,k} - \mu_{j_h,k} \leqslant \varepsilon, Y_{j_h,k} - \mu_{j_h,k} < -T - \mu_{j_h,k}) = 0,$$

and, hence, for $J > J_0$

$$R_J = \sup_{f\in\mathrm{Lip}(\gamma)} \mathsf{P}\left(\max_{j,k}\left|\hat{Y}_{j,k} - \mu_{j,k}\right| > \varepsilon\right) = 1.$$

This fact shows that the hard thresholding makes meaningless the estimation of the loss in the sense of (4). It is an interesting observation, since this effect is not observed when estimating the mean-square error.

4. Soft thresholding

Let us now consider the estimates of the wavelet coefficients obtained using the soft threshold processing: $\hat{Y}_{j,k} = \rho_T^{(s)}(Y_{j,k})$.

Let the function $g_1(J) > 0$ arbitrarily slowly tend to zero, and the function $g_2(J) > 0$ arbitrarily slowly unboundedly increase in J.

Let the indices j_1 and j_2 ($j_1 < j_2$) be such that

$$|\mu_{j,k}| \leqslant (g_1(J))^{-(\gamma+1/2)}, \quad j_1 \leqslant j \leqslant j_2 - 1;$$
$$|\mu_{j,k}| \leqslant (g_2(J))^{-(\gamma+1/2)}, \quad j_2 \leqslant j \leqslant J - 1.$$

By virtue of (1)

$$j_i = \frac{J}{2\gamma+1} + \log_2 g_i(J), \quad i = 1, 2. \tag{5}$$

We split (3) into three sums:

$$\sum_{j=0}^{J-1}\sum_{k=0}^{2^j-1}\left|\ln\mathsf{P}\left(\left|\hat{Y}_{j,k} - \mu_{j,k}\right| \leqslant \varepsilon\right)\right| = \sum_{j=0}^{j_1-1}\sum_{k=0}^{2^j-1}\left|\ln\mathsf{P}\left(\left|\hat{Y}_{j,k} - \mu_{j,k}\right| \leqslant \varepsilon\right)\right| +$$
$$+ \sum_{j=j_1}^{j_2-1}\sum_{k=0}^{2^j-1}\left|\ln\mathsf{P}\left(\left|\hat{Y}_{j,k} - \mu_{j,k}\right| \leqslant \varepsilon\right)\right| + \sum_{j=j_2}^{J-1}\sum_{k=0}^{2^j-1}\left|\ln\mathsf{P}\left(\left|\hat{Y}_{j,k} - \mu_{j,k}\right| \leqslant \varepsilon\right)\right| \equiv$$
$$\equiv S_1 + S_2 + S_3. \tag{6}$$

Consider S_3. Note that for any fixed $\varepsilon > 0$ there exists J_1 such that $(g_2(J))^{-(\gamma+1/2)} \leqslant \varepsilon$ for all $J > J_1$. Thus, $|\mu_{j,k}| \leqslant \varepsilon$ for $j_2 \leqslant j \leqslant J-1$. Consequently, for each term from S_3 we have for $J > J_1$

$$\ln\mathsf{P}\left(\left|\hat{Y}_{j,k}-\mu_{j,k}\right| \leqslant \varepsilon\right) = \ln\Big[\mathsf{P}(|\mu_{j,k}| \leqslant \varepsilon, |Y_{j,k}| \leqslant T)+$$

$$+\mathsf{P}(|Y_{j,k}-\mu_{j,k}-T| \leqslant \varepsilon, Y_{j,k} > T) + \mathsf{P}(|Y_{j,k}-\mu_{j,k}+T| \leqslant \varepsilon, Y_{j,k} < -T)\Big] =$$

$$= \ln\left[1-2H(T+\varepsilon)\right].$$

Since

$$\ln(1-x) \sim -x \quad \text{when} \quad x \to 0,$$

we conclude that

$$S_3 \asymp 2^J H(T+\varepsilon). \tag{7}$$

To find the lower estimate of the optimal threshold, we note that the loss function (3) tends to infinity the faster the more signal samples satisfy $|\mu_{j,k}| > \varepsilon$. According to the definition (5), the maximum number of such samples is of the order of 2^{j_2}. Assuming that for all terms of the sums S_1 and S_2 from (6) the relation $|\mu_{j,k}| > \varepsilon$ is satisfied, we get

$$S_1+S_2 \asymp \sum_{j=0}^{j_2-1}\sum_{k=0}^{2^j-1}\big|\ln\big[\mathsf{P}(T-\varepsilon \leqslant Y_{j,k}-\mu_{j,k} \leqslant T+\varepsilon, Y_{j,k} > T)+$$

$$+\mathsf{P}(-T-\varepsilon \leqslant Y_{j,k}-\mu_{j,k} \leqslant -T+\varepsilon, Y_{j,k} < -T)\big]\big| =$$

$$= \sum_{j=0}^{j_2-1}\sum_{k=0}^{2^j-1}\left|\ln\left[H(T-\varepsilon)-H(T+\varepsilon)\right]\right| \asymp$$

$$\asymp 2^{\frac{J}{2\gamma+1}} g_2(J)\left|\ln\left[H(T-\varepsilon)-H(T+\varepsilon)\right]\right|.$$

Let us equate the orders of S_1+S_2 and S_3. The threshold $T_m^{(h)}$, satisfying the relation

$$\frac{H(T+\varepsilon)}{\left|\ln\left[H(T-\varepsilon)-H(T+\varepsilon)\right]\right|} \asymp 2^{-\frac{2\gamma J}{2\gamma+1}} g_2(J), \tag{8}$$

ensures equality of orders and, thus, is an asymptotic lower bound for the threshold that is optimal in the sense of the loss function R_J.

Now let us find the upper bound for the optimal threshold. Note that for a constant C_f from the inequality (1) there is a function $f \in \mathrm{Lip}(\gamma, L)$ such that this inequality turns to equality for $0 \leqslant j \leqslant j_1-1$ (see [7]). Therefore, since T increases, there is $J_2 > 0$ such that for all $\varepsilon > 0$ and $J > J_2$, the inequality $|\mu_{j,k}| > \varepsilon$ holds for $0 \leqslant j \leqslant j_1-1$. Using the above argument, we get

$$S_1 \asymp 2^{\frac{J}{2\gamma+1}} g_1(J)\left|\ln\left[H(T-\varepsilon)-H(T+\varepsilon)\right]\right|.$$

Let us equate the orders of S_1 and S_3. The threshold $T_M^{(h)}$, satisfying the relation

$$\frac{H(T+\varepsilon)}{\left|\ln\left[H(T-\varepsilon)-H(T+\varepsilon)\right]\right|} \asymp 2^{-\frac{2\gamma J}{2\gamma+1}} g_1(J), \tag{9}$$

ensures equality of orders.

Note that the sum S_2 is not present in these arguments. This means that the true value of the asymptotically optimal threshold T must be no greater than $T_M^{(h)}$.

The above considerations allow us to formulate the following statement.

Theorem. *For an optimal soft threshold value that minimizes the rate at which the loss function* (4) *tends to infinity, the following inequality holds, starting with some* J :

$$T_m^{(s)} \leqslant T \leqslant T_M^{(s)},$$

where $T_m^{(s)}$ *and* $T_M^{(s)}$ *are defined by the relations* (8) *and* (9) *respectively.*

5. Examples

Consider the class of noise distributions of the form

$$H(x) \asymp x^{\alpha} e^{-\theta x^{\beta}}, \quad \alpha \in \mathbf{R}, \quad \theta \geqslant 0, \quad \beta > 0$$

(it is assumed that $H(x)$ satisfies the requirements listed in Section 2). This class is quite wide and includes distributions with various tails.

Let $\theta = 0$, i.e., the tail of the noise distribution decays in a power-law manner. The thresholds $T_M^{(s)}$ and $T_m^{(s)}$ are found from the relations

$$\frac{(T+\varepsilon)^{\alpha}}{|\ln\left[(T-\varepsilon)^{\alpha}-(T+\varepsilon)^{\alpha}\right]|} \asymp 2^{-\frac{2\gamma J}{2\gamma+1}} g_i(J)$$

for $i = 1, 2$, respectively. Given that $\alpha < 0$ and

$$\ln\left[\frac{(T-\varepsilon)^{\alpha}}{(T+\varepsilon)^{\alpha}} - 1\right] \asymp \ln\frac{2\varepsilon}{T+\varepsilon} \asymp -\ln T,$$

we get for $T_M^{(s)}$ and $T_m^{(s)}$ (for $i = 1, 2$)

$$T^{\alpha} \asymp 2^{-\frac{2\gamma J}{2\gamma+1}} g_i(J) \ln T.$$

Hence,

$$T_M^{(s)} \asymp 2^{\frac{2\gamma J}{|\alpha|(2\gamma+1)}} J (g_1(J))^{-1/|\alpha|}, \qquad T_m^{(s)} \asymp 2^{\frac{2\gamma J}{|\alpha|(2\gamma+1)}} J (g_2(J))^{-1/|\alpha|}.$$

This means that it is not possible to accurately determine the order of T using the described method, since estimates are given with the use of arbitrarily slowly decreasing and increasing functions. In this case, if

$$|\alpha| < \frac{2\gamma}{\gamma + 1/2},$$

then as seen from the inequality (1), all the useful signal is lost during the threshold processing.

Consider the case $\theta \neq 0$. The thresholds $T_M^{(s)}$ and $T_m^{(s)}$ are found from the relations

$$\frac{(T+\varepsilon)^{\alpha} e^{-\theta(T+\varepsilon)^{\beta}}}{\left|\ln\left[(T-\varepsilon)^{\alpha} e^{-\theta(T-\varepsilon)^{\beta}} - (T+\varepsilon)^{\alpha} e^{-\theta(T+\varepsilon)^{\beta}}\right]\right|} \asymp 2^{-\frac{2\gamma J}{2\gamma+1}} g_i(J), \quad i = 1, 2.$$

Note that $(T \pm \varepsilon)^{\alpha} \asymp T^{\alpha}$,

$$\ln\left[e^{-\theta(T-\varepsilon)^{\beta}} - e^{-\theta(T+\varepsilon)^{\beta}}\right] \asymp -\theta T^{\beta} + (\beta - 1)\ln T \ \text{ for } 0 < \beta < 1,$$

$$\ln\left[e^{-\theta(T-\varepsilon)^{\beta} + \theta(T+\varepsilon)^{\beta}} - 1\right] \asymp C \ \text{ for } \beta = 1,$$

$$\ln\left[e^{-\theta(T-\varepsilon)^\beta+\theta(T+\varepsilon)^\beta}-1\right]\asymp-\theta(T-\varepsilon)^\beta+\theta(T+\varepsilon)^\beta \quad \text{for } \beta>1,$$

i.e., for all $\beta>0$, we have

$$\left|\ln\left[(T-\varepsilon)^\alpha e^{-\theta(T-\varepsilon)^\beta}-(T+\varepsilon)^\alpha e^{-\theta(T+\varepsilon)^\beta}\right]\right|\asymp T^\beta.$$

Therefore, for $\theta\neq0$, the orders of the estimates of $T_m^{(h)}$ and $T_M^{(h)}$ are the same. Thus, when $\theta\neq0$, for the asymptotically optimal soft threshold we get the relation

$$T\sim\left(\frac{2\gamma\ln2^J}{\theta(2\gamma+1)}\right)^{1/\beta}.$$

In particular, for $\alpha=-1$, $\theta=1/(2\sigma^2)$, $\beta=2$, the noise has a centered normal distribution with a variance σ^2, for which, as shown in [10], the asymptotically optimal soft threshold is of the order

$$T\sim\sigma\sqrt{\frac{4\gamma\cdot\ln2^J}{2\gamma+1}}.$$

Acknowledgments

This research is partly supported by the Russian Foundation for Basic Research (project No. 19–07–00352).

REFERENCES

1. D. Donoho and I. Johnstone, “Adapting to unknown smoothness via wavelet shrinkage,” *J. Am. Stat. Assoc.*, **90**, 1200–1224 (1995).
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, Berlin (2001).
4. J. Sadasivan, S. Mukherjee, and C. S. Seelamantula, “An optimum shrinkage estimator based on minimum-probability-of-error criterion and application to signal denoising,” in: *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, Piscataway, NJ (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).
8. A.V. Markin and O.V.Shestakov, “Consistency of risk estimation with thresholding of wavelet coefficients,” *Moscow Univ. Comput. Math. Cybern.*, **34**, No. 1, 22–30 (2010).
9. O.V. Shestakov, “Asymptotic normality of adaptive wavelet thresholding risk estimation,” *Dokl. Math.*, **86**, No. 1, 556–558 (2012).
10. A. Kudryavtsev and O. Shestakov, “The asymptotic behavior of the optimal threshold minimizing the probability-of-error criterion,” *J. Math. Sci.*, **234**, No. 6, 810–815 (2018).