ISSN 1064-5624, Doklady Mathematics, 2016, Vol. 94, No. 3, pp. 615–619. © Pleiades Publishing, Ltd., 2016.
Original Russian Text © A.A. Kudryavtsev, O.V. Shestakov, 2016, published in Doklady Akademii Nauk, 2016, Vol. 471, No. 1, pp. 11–15.

MATHEMATICS

Asymptotically Optimal Wavelet Thresholding in Models with Non-Gaussian Noise Distributions

A. A. Kudryavtsev[a]* and O. V. Shestakov[a,b]**

Presented by Academician of the RAS I.A. Sokolov June 2, 2016

Received June 2, 2016

Abstract—The problem of nonparametric estimation of a signal function by thresholding the coefficients of its wavelet decomposition is considered. In models with various noise distributions, asymptotically optimal thresholds and orders of the loss functions are calculated on the basis of probabilities of errors in the calculation of wavelet coefficients.

DOI: 10.1134/S1064562416060028

When methods of wavelet analysis are used for nonparametric estimation of signals from noisy observations, it is usually assumed that the signal function belongs to a certain class of functions. This assumption leads to the conclusion that the signal function admits an "economical" representation in the space of its wavelet coefficients, i.e., the majority of the wavelet coefficients are small in absolute value and can be neglected. Such considerations lead to so-called thresholding procedures, which set all wavelet coefficients whose absolute values are smaller than a certain threshold to zero. In models with an additive Gaussian noise, the statistical properties of these procedures have been well studied, and expressions for "optimal" thresholds oriented to various loss functions were obtained (see, e.g., [1–4]). This paper considers a model of wavelet coefficients of an additive-noise signal function whose distribution is subject to very general constraints. For the class of Lipschitz continuous signals, we give relations that make it possible to calculate a threshold ensuring an asymptotically optimal order of the loss function, which is based on the probability that the error of the computation of wavelet coefficients exceeds a certain critical level. We also exemplify the calculation of an asymptotically optimal threshold for various distributions of noise coefficients.

[a] *Faculty of Computational Mathematics and Cybernetics, Moscow State University, Moscow, 119991 Russia*
[b] *Institute of Informatics Problems, Federal Research Center "Computer Science and Control," Russian Academy of Sciences, Moscow, 119933 Russia*
**e-mail: nubigena@mail.ru*
***e-mail: oshestakov@cs.msu.su*

1. A MODEL FOR WAVELET COEFFICIENTS AND A METHOD FOR SUPPRESSING NOISE

Consider the class of functions $f \in L^2(\mathbf{R})$ given on a finite interval $[a, b]$ and uniformly Lipschitz regular with an exponent $\gamma > 0$: $f \in \mathrm{Lip}(\gamma)$. The signal function f is usually given at discrete sampling positions. Suppose that the number of these positions equals 2^J for some $J > 0$. A discrete wavelet transform multiplies the vector of values of f by the orthogonal matrix determined by some wavelet function ψ (see [5]). This yields a set $\{\mu_{j,k}\}_{j=0,\ldots,J-1,\,k=0,\ldots,2^j-1}$ of wavelet coefficients; the subscript j is referred to as the scale and the subscript k, as the shift.

If a wavelet function is M-times continuously differentiable ($M \ge \gamma$), has M zero moments, and rapidly decreases at infinity together with its derivatives, i.e., for all $0 \le k \le M$ and any $m \in \mathbf{N}$, there exists a constant C_m such that

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

for all $x \in \mathbf{R}$, then there exists a $C_f > 0$ such that [5]

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

In what follows, we assume that the wavelet function ψ meets these requirements.

In real-life observations, there is a noise, which we assume to be additive. In this paper, we use the following model for empirical wavelet coefficients:

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

where the $\mu_{j,k}$ are the discrete wavelet coefficients of the "pure" signal f and the $W_{j,k}$ are "noise" coefficients, which are assumed to be independent and have symmetric absolutely continuous differentiable distribution function $\mathrm{P}(W_{j,k} < x) = 1 - H(x)$.

Suppose that $0 < H(x) < 1$ and let $h(x)$ denote the derivative (density) of the distribution function of $W_{j,k}$. Suppose also that $h(x)$ has no discontinuities of the second kind.

The noise is usually suppressed by thresholding; in essence, this is setting the coefficients with absolute values not exceeding a certain threshold to zero.

Let $\hat{Y}_{j,k}$ denote the estimate of the corresponding wavelet coefficient obtained by means of thresholding, which is determined by a function T for the threshold $\rho_T(x)$: $\hat{Y}_{j,k} = \rho_T(Y_{j,k})$. This paper considers hard thresholding functions $\rho_T^{(h)}(x) = x \cdot 1(|x| > T)$ and soft thresholding functions $\rho_T^{(s)}(x) = \mathrm{sgn}(x)(|x| - T)_+$.

Consider a loss function based on the probability that the error in calculating the wavelet coefficients exceeds a certain critical level $\varepsilon > 0$. For this purpose, we take a two-dimensional random variable (ξ, η) not depending on all $W_{j,k}$ and having discrete uniform distribution on the set of indices $j = 0,1,\ldots,J-1$, $k = 0,1,\ldots,2^j-1$ and set

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

Such a loss function is the averaged probability that the error of the calculation of wavelet coefficients exceeds ε. This definition of a loss function generalizes that proposed in [4]. In the same paper [4], it was shown that estimates aimed at minimizing the loss $r_J(f)$ give results comparable with and sometimes even better than those obtained by using estimates minimizing the standard deviation.

The purpose of this paper is to find an asymptotically optimal threshold for processing an observed signal in the class of functions $f \in \mathrm{Lip}(\gamma)$, for which the loss function is defined as

$$R_J = \sup_{f \in \mathrm{Lip}(\gamma)} r_J(f), \tag{2}$$

i.e., a threshold asymptotically optimal in the minimax sense. A detailed study of the behavior of an asymptotically optimal threshold for standard deviation in a model with an additive Gaussian noise can be found in [2, 3]. In [1], a method for determining an adaptive optimal threshold was also proposed, which can be used to estimate the thresholding risk for a particular function. This method is based on constructing an unbiased risk estimator, whose statistical properties were thoroughly studied in [6, 7]. Note that a "reasonable" threshold must increase with J (see [3]). However, to simplify formulas, we do not explicitly specify the dependence of the threshold on J in what follows.

We use $\asymp$ to denote the order of a quantity under consideration with respect to J; i.e., $a_J \asymp b_J$ if $C_1 \cdot b_J \le a_J \le C_2 \cdot b_J$ for some positive constants C_1 and C_2, provided that J is sufficiently large. By $a_J \sim b_J$ we mean that $\lim\limits_{J\to\infty}\frac{a_J}{b_J} = 1$.

2. HARD THRESHOLDING

Let $\hat{Y}_{j,k} = \rho_T^{(h)}(Y_{j,k})$.

Suppose that a function $g_1(J) > 0$ arbitrarily slowly increases without bound and a function $d(J) > 0$ arbitrarily slowly tends to zero with increasing J. We set

$$g_2(J) = \min\left\{d(J), \inf_{x \in [T-d(J), T+d(J)]}\left\{C \cdot \frac{H(x)}{h(x)}\right\}\right\},$$

where C is a positive constant.

Inequality (1) makes it possible to divide the whole set $\{0,1,\ldots,J-1\}$ of indices into three classes, depending on $|\mu_{j,k}|$. Suppose that indices j_1 and j_2 $(j_1 < j_2)$ are such that

$$\begin{aligned} |\mu_{j,k}| &\le g_1(J), \quad j_1 \le j \le j_2 - 1; \\ |\mu_{j,k}| &\le g_2(J), \quad j_2 \le j \le J - 1. \end{aligned} \tag{3}$$

By virtue of (1), we have

$$j_i = \frac{J}{2\gamma+1} - \frac{1}{\gamma+1/2}\log_2 g_i(J), \quad i = 1,2. \tag{4}$$

Let us split the sum in the definition of the loss function into three components:

$$\begin{aligned} \sum_{j=0}^{J-1}\sum_{k=0}^{2^j-1}\mathrm{P}\left(\left|\hat{Y}_{j,k} - \mu_{j,k}\right| > \varepsilon\right) &= \sum_{j=0}^{j_1-1}\sum_{k=0}^{2^j-1}\mathrm{P}\left(\left|\hat{Y}_{j,k} - \mu_{j,k}\right| > \varepsilon\right) \\ &+ \sum_{j=j_1}^{j_2-1}\sum_{k=0}^{2^j-1}\mathrm{P}\left(\left|\hat{Y}_{j,k} - \mu_{j,k}\right| > \varepsilon\right) \\ &+ \sum_{j=j_2}^{J-1}\sum_{k=0}^{2^j-1}\mathrm{P}\left(\left|\hat{Y}_{j,k} - \mu_{j,k}\right| > \varepsilon\right) \equiv S_1 + S_2 + S_3. \end{aligned} \tag{5}$$

Consider S_3. Note that, for any $\varepsilon > 0$, there exists a $J_0 = J_0(\varepsilon)$ such that $g_2(J) \le \varepsilon$ and $\varepsilon \le cT$, $0 < c < 1$, for all $J > J_0$, and there exists a $J_1 = J_1(\varepsilon, c) \ge J_0$ such that, for any $J > J_1$, we have

$$\mu_{j,k} + \varepsilon \ge 0, \quad \mu_{j,k} - \varepsilon \le 0,$$
$$\mu_{j,k} + \varepsilon \le T, \quad \mu_{j,k} - \varepsilon \ge -T$$

for $j_2 \le j \le J-1$. Therefore, for one summand in S_3, we can write

$$\begin{aligned} &\mathrm{P}\left(\left|\hat{Y}_{j,k} - \mu_{j,k}\right| > \varepsilon\right) = \mathrm{P}\left(\left|\rho_T^{(h)}(Y_{j,k}) - \mu_{j,k}\right| > \varepsilon\right) \\ &= \mathrm{P}\left(\left|Y_{j,k}\right| > T\right) = H(T - \mu_{j,k}) + H(T + \mu_{j,k}), \end{aligned} \tag{6}$$

provided that $J > J_1$.

It can be shown that condition (3) implies

$$H(T \pm \mu_{j,k}) \asymp H(T)$$

therefore, for $j_2 \le j \le J-1$, the quantities $|\mu_{j,k}|$ do not affect the order of the right-hand side of (6). Taking into account the fact that the number of terms in S_3 is of order 2^J, we obtain

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

Let us find an upper bound for the loss function (2) under hard thresholding.

For this purpose, suppose that all summands of S_1 and S_2 in (5) are bounded away from zero by some constant. Then (4) implies

$$S_1 + S_2 \asymp 2^{\frac{J}{2\gamma+1}} (g_2(J))^{-\frac{1}{\gamma+1/2}}. \tag{8}$$

A threshold $T_m^{(h)}$ satisfying the relation

$$H(T)(g_2(J))^{\frac{1}{\gamma+1/2}} \asymp 2^{-\frac{2\gamma J}{2\gamma+1}}, \tag{9}$$

ensures the equality of the orders of the right-hand sides of (7) and (8) and, therefore, is a lower bound for a threshold asymptotically optimal in the sense of the loss function R_J.

Now, let us find a lower bound for the loss function (2). Note, that for any constant in (1), there exists a function $f \in \mathrm{Lip}(\gamma)$ such that (1) becomes an equality for $0 \le j \le j_1 - 1$ (see [5]). Therefore, for any $\varepsilon > 0$, there exists a function $f \in \mathrm{Lip}(\gamma)$ such that $|\mu_{j,k}| > \varepsilon$ for $0 \le j \le j_1 - 1$. Thus,

$$\mathrm{P}\left(\left|\hat{Y}_{j,k} - \mu_{j,k}\right| > \varepsilon\right) \ge \mathrm{P}(\left|Y_{j,k} - \mu_{j,k}\right| > \varepsilon) = 2H(\varepsilon).$$

In this case, the order of the sum S_1 in (5) equals the number of summands, i.e.,

$$S_1 \asymp 2^{j_1} = 2^{\frac{J}{2\gamma+1}} (g_1(J))^{-\frac{1}{\gamma+1/2}}. \tag{10}$$

Let us equate the orders of S_1 and S_3 in (10) and (7). The equality of orders in the case under consideration ensures that the threshold $T_M^{(h)}$ satisfies the relation

$$H(T) \asymp 2^{-\frac{2\gamma J}{2\gamma+1}} (g_1(J))^{-\frac{1}{\gamma+1/2}}. \tag{11}$$

Note that the above argument does not use the sum S_2. This means that the order of the true value R_J is no lower than the given one, i.e., the order under consideration is a lower bound for the true order of the loss function, and $T_M^{(h)}$ is an upper bound for an asymptotically optimal threshold T.

These considerations allow us to state the following theorem.

Theorem 1. *For hard thresholding with asymptotically optimal threshold, the loss function* (2) *satisfies the inequalities*

$$\begin{aligned} &C_m^{(h)} \cdot 2^{-\frac{2\gamma}{2\gamma+1}J} (g_1(J))^{-\frac{1}{\gamma+1/2}} \\ &\le R_J \le C_M^{(h)} \cdot 2^{-\frac{2\gamma}{2\gamma+1}J} (g_2(J))^{-\frac{1}{\gamma+1/2}}, \end{aligned}$$

where $C_m^{(h)}$ and $C_M^{(h)}$ are positive constants. For hard thresholding with asymptotically optimal threshold minimizing the order of the loss function (2)*, the following inequality holds for sufficiently large J*:

$$T_m^{(h)} \le T \le T_M^{(h)},$$

where $T_m^{(h)}$ and $T_M^{(h)}$ are determined by (9) *and* (11)*, respectively.*

3. SOFT THRESHOLDING

Let $\hat{Y}_{j,k} = \rho_T^{(s)}(Y_{j,k})$. Suppose that a function $g_1(J) > 0$ arbitrarily slowly increases without bound and a function $g_2(J) > 0$ arbitrarily slowly tends to zero with increasing J. Unlike in the preceding section, the decrease of the function $g_2(J)$ does not depend on the behavior of the tail of the noise distribution.

As in Section 2, we break the index set $\{0, 1, \ldots, J-1\}$ into three classes by using j_1 and j_2 in (4) and the sum in the definition of the loss function, into the three components (5).

Consider S_3. Take a positive number ε. Note that, for any $\varepsilon > 0$, there exists a $J_0 = J_0(\varepsilon)$ such that $g_2(J) \le \varepsilon$ for all $J > J_0$, i.e., $|\mu_{j,k}| \le \varepsilon$ for $j_2 \le j \le J-1$. Therefore, for one summand of S_3, we have

$$\begin{aligned} &\mathrm{P}\left(\left|\hat{Y}_{j,k} - \mu_{j,k}\right| > \varepsilon\right) = \mathrm{P}\left(\left|\rho_T^{(s)}(Y_{j,k}) - \mu_{j,k}\right| > \varepsilon\right) \\ &= \mathrm{P}(\left|Y_{j,k} - \mu_{j,k}\right| > T + \varepsilon) = 2H(T + \varepsilon), \end{aligned}$$

provided that $J > J_0$. Thus,

$$S_3 \asymp 2^J H(T+\varepsilon).$$

Assuming that all summands of S_1 and S_2 in (5) are bounded away from zero by some constant, we obtain $S_1 + S_2 \asymp 2^{j_2}$. For any constant C_f in (1), there exists a function $f \in \mathrm{Lip}(\gamma)$ such that inequality (1) becomes an equality for $0 \le j \le j_1 - 1$; therefore, for each $\varepsilon > 0$, there exists a function $f \in \mathrm{Lip}(\gamma)$ such that $|\mu_{j,k}| > \varepsilon$ for $0 \le j \le j_1 - 1$. Thus, for any $\delta \in (0,1)$ and sufficiently large J, we have

$$\mathrm{P}\left(\left|\hat{Y}_{j,k} - \mu_{j,k}\right| > \varepsilon\right) = 1 - H(T-\varepsilon) + H(T+\varepsilon) \ge \delta.$$

This implies $S_1 \asymp 2^{j_1}$.

Repeating the argument in Section 2, we see that the upper bound $T_M^{(s)}$ and the lower bound $T_m^{(s)}$ of an asymptotically optimal threshold, which give, respectively, lower and upper bounds for the loss function (2), must satisfy the relation

$$H(T+\varepsilon) \asymp 2^{-\frac{2\gamma J}{2\gamma+1}} (g_i(J))^{-\frac{1}{\gamma+1/2}} \tag{12}$$

for $i = 1, 2$, respectively.

Thus, the following theorem is valid.

Theorem 2. *For soft thresholding with asymptotically optimal threshold, the loss function* (2) *satisfies the inequalities*

$$C_m^{(s)} \cdot 2^{-\frac{2\gamma}{2\gamma+1}J} (g_1(J))^{-\frac{1}{\gamma+1/2}}$$
$$\le R_J \le C_M^{(s)} \cdot 2^{-\frac{2\gamma}{2\gamma+1}J} (g_2(J))^{-\frac{1}{\gamma+1/2}},$$

where $C_m^{(s)}$ and $C_M^{(s)}$ are some positive constants. For soft thresholding with asymptotically optimal threshold minimizing the order of the loss function (2), *the following inequality holds for sufficiently large J*:

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

where $T_M^{(s)}$ and $T_m^{(s)}$ satisfy relation (12) *for $i = 1, 2$, respectively.*

4. EXAMPLES

Consider a noise distribution of the form

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

(it is assumed that $H(x)$ satisfies all requirements listed in Section 1). Since

$$\frac{H(x)}{h(x)} \asymp \frac{x}{\beta\theta x^{\beta} - \alpha},$$

it follows that, under hard thresholding, we obtain $g_2(J) \asymp T^{-\beta+1}$ for $\theta \ne 0$ and $\beta > 1$ and $g_2(J) = d(J)$ otherwise. For $\theta \ne 0$, we have $T_m^{(h)} \sim T_M^{(h)}$; therefore, for an asymptotically optimal hard threshold,

$$T \sim \left(\frac{\gamma \ln 2^J}{\theta(\gamma+1/2)}\right)^{1/\beta} \tag{13}$$

for $\theta \ne 0$ and $\beta > 1$. In particular, for $\alpha = -1$, $\theta = \frac{1}{2\sigma^2}$, and $\beta = 2$, the noise has a centered normal distribution with variance σ^2, for which, as shown in [8], an asymptotically optimal hard threshold has order

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

Moreover, it is easy to show that relation (13) remains valid for $\theta \ne 0$ and $0 < \beta \le 1$.

Consider the case $\theta = 0$, in which the tail of the noise distribution decreases polynomially (provided that $\alpha < 0$). Asymptotic estimates for an optimal hard threshold have the form

$$T_m^{(h)} \asymp (d(J))^{\frac{1}{|\alpha|(\gamma+1/2)}} \cdot 2^{\frac{\gamma J}{|\alpha|(\gamma+1/2)}},$$
$$T_M^{(h)} \asymp (g_1(J))^{\frac{1}{|\alpha|(\gamma+1/2)}} \cdot 2^{\frac{\gamma J}{|\alpha|(\gamma+1/2)}}.$$

This means that it is impossible to determine the exact order of T by the method described above, because the bounds are specified by using arbitrarily slowly decreasing and increasing functions. Moreover, if

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

then, as seen from (1), the whole useful signal is lost under thresholding.

Under soft thresholding, bounds for an asymptotically optimal threshold satisfy relation (12), which, unlike in relations (9) and (11), involves the tail of the noise distribution at the point $T + \varepsilon$ instead of $H(T)$. Since $T + \varepsilon \sim T$, it follows that all relations for asymptotically optimal thresholds mentioned in this section are also valid for soft thresholding.

ACKNOWLEDGMENTS

This work was supported by the Russian Science Foundation, project no. 14-11-00364.

REFERENCES

1. D. Donoho and I. Johnstone, J. Am. Stat. Assoc. **90**, 1200–1224 (1995).

2. D. Donoho and I. M. Johnstone, Ann. Stat. **26** (3), 879–921 (1998).
3. M. Jansen, *Noise Reduction by Wavelet Thresholding* (Springer, New York, 2001).
4. J. Sadasivan, S. Mukherjee, and C. S. Seelamantula, in *Proceedings of the 39th IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Florence, Italy, 2014* (Florence, 2014), pp. 4249–4253.
5. S. Mallat, *A Wavelet Tour of Signal Processing* (Academic, New York, 1999).
6. A. V. Markin and O. V. Shestakov, Moscow Univ. Comput. Math. Cybernet. **34** (1), 22–30 (2010).
7. O. V. Shestakov, Dokl. Math. **86** (1), 556–558 (2012).
8. A. A. Kudryavtsev and O. V. Shestakov, Dokl. Math. **93** (3), 295–299 (2016).

Translated by O. Sipacheva