DOI 10.1007/s10958-020-04854-6
Journal of Mathematical Sciences, Vol. 248, No. 1, July, 2020

MEAN-SQUARE RISK OF THE THRESHOLD PROCESSING IN THE PROBLEM OF INVERTING THE RADON TRANSFORM WITH A RANDOM SAMPLE SIZE

O. V. Shestakov[1]

Methods for reconstructing tomographic images based on the inversion of the Radon transform are used in problems arising in medicine, biology, astronomy, and many other fields. In the presence of noise in the projection data, as a rule, it is necessary to apply regularization methods. Recently, methods of the threshold processing of wavelet expansion coefficients have become popular. The analysis of errors of these methods is an important practical task, since it makes it possible to assess the quality of both the methods themselves and the equipment used. When using threshold processing, it is usually assumed that the number of expansion coefficients is fixed, and the noise distribution is Gaussian. This model is well studied in the literature, and the optimal values of the threshold processing parameters are calculated for different classes of functions. However, in some situations, the amount of data is not known in advance and has to be modeled with a certain random variable. In this paper, we consider a model with a random amount of data containing Gaussian noise, and estimate the order of the mean-square risk with an increasing number of decomposition coefficients.

1. Introduction

Problems of tomographic image reconstruction arise in medicine, biology, astronomy, and in many other areas. In this case, as a rule, mathematical models based on the Radon transform are used. The problem of inverting this transform is ill-posed, and if there is a noise in the projection data, the wavelet analysis methods based on the threshold processing of the wavelet coefficients are applicable for its regularization [1, 2]. The papers [3–5] analyze the mean-square risk and calculate the optimal values of the parameters of such regularization methods.

In some cases, the amount of data available for analysis is not known in advance. Such situations may arise, for example, in the case of missing data or lack of information about the characteristics of the equipment used. In such a case, it is assumed that the data sample size is a random variable with some specified distribution. In this paper, we consider a model with a random number of expansion coefficients of a function that describes the projection data of a tomographic image “polluted” with white Gaussian noise, and estimate the order of the mean-square risk of the threshold processing method. Similar results in the model of direct observation of a signal were obtained in [6, 7].

2. Inversion of the Radon transform

The mathematical model of the problem of tomographic image reconstruction is based on the so-called Radon transform. Denote by $\mathrm{Lip}(\gamma, L)$ the class of uniformly Lipschitz-regular functions, where $\gamma > 0$ is a Lipschitz exponent and $L > 0$ is a Lipschitz constant [8]. Let the image be described by the function $f(x, y) \in \mathrm{Lip}(\gamma, L)$ with a compact support (without loss of generality, we assume that this is a circle of unit radius with the center at the origin) and uniformly Lipschitz-regular with some exponent

[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 XXXV International Seminar on Stability Problems for Stochastic Models, Perm, Russia, September 24–28, 2018. Part II.

 1072-3374/20/2481-046

$\gamma > 0$. The Radon transform of a function f is

$$Rf(s,\theta) = \int\limits_{L_{s,\theta}} f(x,y)dl,$$

where

$$L_{s,\theta} = \big\{(x,y)\colon x\cos\theta + y\sin\theta - s = 0\big\}.$$

One of the possible methods for inverting the transform Rf is the wavelet-vaguelette decomposition [1].

Let $\phi(x)$ and $\psi(x)$ be the scaling function and the wavelet function, respectively. Define

$$\begin{aligned}\psi^{[1]}_{j,k_1,k_2}(x,y) &= 2^j\phi(2^jx-k_1)\psi(2^jy-k_2),\\ \psi^{[2]}_{j,k_1,k_2}(x,y) &= 2^j\psi(2^jx-k_1)\phi(2^jy-k_2),\\ \psi^{[3]}_{j,k_1,k_2}(x,y) &= 2^j\psi(2^jx-k_1)\psi(2^jy-k_2).\end{aligned} \tag{1}$$

The family $\{\psi^{[\lambda]}_{j,k_1,k_2}\}$ forms an orthonormal basis in $L^2(\mathbb{R}^2)$. The index j in (1) is called the scale, and the indices k_1, k_2 are called the shifts.

The wavelet decomposition of the function f has the form

$$f = \sum_{\lambda,j,k_1,k_2} \langle f, \psi^{[\lambda]}_{j,k_1,k_2}\rangle \psi^{[\lambda]}_{j,k_1,k_2}. \tag{2}$$

If the wavelet function has M vanishing moments and M continuous derivatives ($M \geqslant \gamma$) and quickly decays at infinity, then the expansion coefficients in (2) satisfy the inequality

$$\left|\langle f, \psi^{[\lambda]}_{j,k_1,k_2}\rangle\right| \leqslant \frac{A}{2^{j(\gamma+1)}}, \tag{3}$$

where A is a positive constant [8].

Let us define the functions $\xi^{[\lambda]}_{j,k_1,k_2}(s,\theta)$:

$$\xi^{[\lambda]}_{j,k_1,k_2}(s,\theta) = \frac{2^{-j/2}}{4\pi} I^{-1}\left[R\psi^{[\lambda]}_{j,k_1,k_2}\right](s,\theta),$$

where I^α is the Riesz potential, defined in the Fourier space by the formula $\widehat{I^\alpha g}(\omega) = |w|^{-\alpha}\widehat{g}(\omega)$. These functions are called the “vaguelettes” [1]. They satisfy the relation

$$\langle f, \psi^{[\lambda]}_{j,k_1,k_2}\rangle = 2^{j/2}\langle Rf, \xi^{[\lambda]}_{j,k_1,k_2}\rangle.$$

The sequence $\{\xi^{[\lambda]}_{j,k_1,k_2}\}$ forms a stable basis [3], and the wavelet-vaguelette decomposition of f has the form

$$f = \sum_{\lambda,j,k_1,k_2} 2^{j/2}\langle Rf, \xi^{[\lambda]}_{j,k_1,k_2}\rangle \psi^{[\lambda]}_{j,k_1,k_2}. \tag{4}$$

The decomposition (4) uses only projection data, and it serves as the basis of the reconstruction method.

3. Data model and de-noising method

In practice, the image is described by discrete samples of a certain function. Moreover, since the support of the function is the unit circle, $(s,\theta)\in[-1,1]\times[0,\pi]$, and the model of noisy projection data considered in this paper looks as follows:

$$Y_{i,j}=Rf(-1+2i/N,j\pi/N)+e_{i,j},\qquad i=1,\dots,N,\ j=1,\dots,N,$$

where $N=2^J$ for some $J>0$. It is assumed that $e_{i,j}$ are independent and have a normal distribution with zero mean and variance σ^2. Then, in the discrete analogue of the wavelet-vaguelette decomposition, the coefficients are described by the model [9]

$$X^{[\lambda]}_{j,k_1,k_2}=\mu^{[\lambda]}_{j,k_1,k_2}+e^{[\lambda]}_{j,k_1,k_2}, \tag{5}$$

where $\mu^{[\lambda]}_{j,k_1,k_2}=2^J\langle Rf,\xi^{[\lambda]}_{j,k_1,k_2}\rangle$, and $e^{[\lambda]}_{j,k_1,k_2}$ have a normal distribution with zero mean and variance σ^2_λ $(\lambda=1,2,3)$ and are no longer independent. The value of σ^2_λ depends on the chosen wavelet basis and λ, but does not depend on k_1, k_2, and j.

To suppress the noise and construct estimates for the coefficients $\mu^{[\lambda]}_{j,k_1,k_2}$, threshold processing of the noisy coefficients of the model (5) is usually used. The meaning of this processing is to remove sufficiently small coefficients, which are considered to be the noise. Estimates of $\mu^{[\lambda]}_{j,k_1,k_2}$ are calculated using the threshold processing function $\rho(x,T_\lambda)$ with some threshold T_λ: $\widehat{\mu}^{[\lambda]}_{j,k_1,k_2}=\rho(X^{[\lambda]}_{j,k_1,k_2},T_\lambda)$. This paper assumes the use of the hard threshold processing function $\rho_H(x,T_\lambda)=x\mathbf{1}(|x|>T_\lambda)$ or the soft threshold processing function $\rho_S(x,T_\lambda)=\mathbf{sgn}(x)\left(|x|-T_\lambda\right)_+$.

The mean-square risk of the threshold processing method is defined by the formula

$$r_J(f)=\frac{1}{2^{2J}}\sum_{j=0}^{J-1}\sum_{k_1=0}^{2^j-1}\sum_{k_2=0}^{2^j-1}\sum_{\lambda=1}^{3}2^j\mathsf{E}\left(\widehat{\mu}^{[\lambda]}_{j,k_1,k_2}-\mu^{[\lambda]}_{j,k_1,k_2}\right)^2. \tag{6}$$

In the case when the number of expansion coefficients is not random, the optimal threshold values and the order of the mean square risk for various classes of signal functions are known. In particular, using the methods from [4,10,11], we can prove the following estimate for the minimax order of the risk (6).

Theorem 1. *When choosing an asymptotically optimal threshold for hard and soft threshold processing, the following relation holds*:

$$\sup_{f\in Lip(\gamma,L)} r_J(f)\leqslant C2^{\frac{-4\gamma}{2\gamma+3}J}J^{\frac{2\gamma+6}{2\gamma+3}},$$

where C is a positive constant.

The asymptotically optimal threshold in Theorem 1 for $J\to\infty$ satisfies the relation

$$T\simeq\sigma\sqrt{\frac{6\gamma+3}{2\gamma+3}\ln 2^{2J}}.$$

In the next section, the order of the mean-square risk of threshold processing in a model with a random number of empirical decomposition coefficients is estimated.

4. Risk for a random number of coefficients

Let M be a positive integer random variable and $N=2^M$. Then the mean-square risk takes the form

$$r(f)=\sum_{J=0}^{\infty}\mathsf{P}\left(N=2^J\right)\frac{1}{2^{2J}}\sum_{j=0}^{J-1}\sum_{k_1=0}^{2^j-1}\sum_{k_2=0}^{2^j-1}\sum_{\lambda=1}^{3}2^j\mathsf{E}\left(\widehat{\mu}^{[\lambda]}_{j,k_1,k_2}-\mu^{[\lambda]}_{j,k_1,k_2}\right)^2, \tag{7}$$

and its asymptotic order depends on the distribution of N. To get meaningful estimates of the order of (7), the value of N must be "large." Consider the sequence N_n, $n = 1, \ldots$, and assume that there is a nonrandom increasing sequence of the 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 distribution) when $n \to \infty$, i.e.,

$$\sup_{x \geqslant 0} |H_n(x) - H(x)| < \frac{\varepsilon_n}{2} \to 0, \qquad n \to \infty, \tag{8}$$

where

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

and $H(x)$ is the limit distribution function. Suppose $H(x)$ does not have an atom at zero. Let us study the behavior of

$$r_n(f) = \sum_{J=0}^{\infty} \mathsf{P}\left(N_n = 2^J\right) \frac{1}{2^{2J}} \sum_{j=0}^{J-1} \sum_{k_1=0}^{2^j-1} \sum_{k_2=0}^{2^j-1} \sum_{\lambda=1}^{3} 2^j \mathsf{E}\left(\widehat{\mu}_{j,k_1,k_2}^{[\lambda]} - \mu_{j,k_1,k_2}^{[\lambda]}\right)^2$$

when $n \to \infty$.

Let $\delta_n \to 0$ and $\alpha_n \to 0$ when $n \to \infty$ so 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

$$\begin{aligned} r_n(f) = &\sum_{J=0}^{[J_n + \log_2 \delta_n]} \mathsf{P}\left(N_n = 2^J\right) \frac{1}{2^{2J}} \sum_{j=0}^{J-1} \sum_{k_1=0}^{2^j-1} \sum_{k_2=0}^{2^j-1} \sum_{\lambda=1}^{3} 2^j \mathsf{E}\left(\widehat{\mu}_{j,k_1,k_2}^{[\lambda]} - \mu_{j,k_1,k_2}^{[\lambda]}\right)^2 + \\ &+ \sum_{J=[J_n + \log_2 \delta_n]+1}^{[J_n - \log_2 \delta_n]} \mathsf{P}\left(N_n = 2^J\right) \frac{1}{2^{2J}} \sum_{j=0}^{J-1} \sum_{k_1=0}^{2^j-1} \sum_{k_2=0}^{2^j-1} \sum_{\lambda=1}^{3} 2^j \mathsf{E}\left(\widehat{\mu}_{j,k_1,k_2}^{[\lambda]} - \mu_{j,k_1,k_2}^{[\lambda]}\right)^2 + \\ &+ \sum_{J=[J_n - \log_2 \delta_n]+1}^{\infty} \mathsf{P}\left(N_n = 2^J\right) \frac{1}{2^{2J}} \sum_{j=0}^{J-1} \sum_{k_1=0}^{2^j-1} \sum_{k_2=0}^{2^j-1} \sum_{\lambda=1}^{3} 2^j \mathsf{E}\left(\widehat{\mu}_{j,k_1,k_2}^{[\lambda]} - \mu_{j,k_1,k_2}^{[\lambda]}\right)^2 \equiv S_1 + S_2 + S_3. \end{aligned}$$

Given (8), for $S_1 + S_3$ we have

$$\begin{aligned} S_1 + S_3 &\leqslant C_0(H_n(\delta_n) + 1 - H_n(\delta_n^{-1}))(J_n + \log_2 \delta_n) 2^{J_n + \log_2 \delta_n} \leqslant \\ &\leqslant C_0(\alpha_n + \varepsilon_n)(J_n + \log_2 \delta_n) 2^{J_n + \log_2 \delta_n} \end{aligned}$$

where C_0 is a positive constant.

For S_2, using Theorem 1, we can obtain an estimate

$$S_2 \leqslant C_1 2^{-\frac{4\gamma}{2\gamma+3}(J_n + \log_2 \delta_n)} (J_n + \log_2 \delta_n)^{\frac{2\gamma+6}{2\gamma+3}}.$$

where C_1 is a positive constant. Thus, the following statement holds.

Theorem 2. *In the model with a random number of empirical coefficients, when choosing an asymptotically optimal threshold, starting with some n the following estimate is valid*:

$$\sup_{f \in Lip(\gamma, L)} r_n(f) \leqslant C_0(\alpha_n + \varepsilon_n)(J_n + \log_2 \delta_n) 2^{J_n + \log_2 \delta_n} + C_1 2^{-\frac{4\gamma}{2\gamma+3}(J_n + \log_2 \delta_n)} (J_n + \log_2 \delta_n)^{\frac{2\gamma+6}{2\gamma+3}}.$$

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

$$T_n \simeq \sigma \sqrt{\frac{6\gamma + 3}{2\gamma + 3} \ln 2^{2(J_n + \log_2 \delta_n)}}.$$

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

$$\sup_{f\in Lip(\gamma)} r_n(f) \leqslant \varepsilon'_n J_n 2^{J_n} + C_2 2^{-\frac{4\gamma}{2\gamma+3}J_n} J_n^{\frac{2\gamma+6}{2\gamma+3}},$$

where ε'_n *characterizes the rate of convergence* $N_n/2^{J_n} \xrightarrow{\mathsf{P}} 1$, *and* C_2 *is a positive constant.*

If ε'_n decreases fast enough, then this estimate coincides with the estimate for the mean-square risk in a model with a non-random number of coefficients.

Corollary 2. *Let* $H(x)$ *be differentiable in a neighborhood of origin, and for some positive constants* b *and* B, *the relation* $b \leqslant H'(x) \leqslant B$ *holds in this neighborhood. Let* $\delta_n \simeq 2^{-\frac{6\gamma+3}{8\gamma+6}J_n}$. *Then, starting with some* n, *the following estimate is valid*:

$$\sup_{f\in Lip(\gamma)} r_n(f) \leqslant C_0 \varepsilon_n J_n 2^{\frac{2\gamma+3}{8\gamma+6}} + C_3 2^{-\frac{2\gamma}{4\gamma+3}J_n} J_n^{\frac{2\gamma+6}{2\gamma+3}}, \tag{9}$$

where C_3 *is a positive constant.*

Thus, the mean-square risk for a random number of empirical coefficients may tend to zero much slower than the mean-square risk for a non-random number of coefficients.

Acknowledgments

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

REFERENCES

1. D. Donoho, “Nonlinear solution of linear inverse problems by wavelet-vaguelette decomposition,” *Appl. Comput. Harmon. Anal.*, **2**, 101–126 (1995).
2. M. Jansen, *Noise Reduction by Wavelet Thresholding*, Springer Verlag (2001).
3. N. Lee, *Wavelet-vaguelette decompositions and homogenous equations*, Ph.D. Thesis, Purdue University, West Lafayette (1997).
4. F. Abramovich and B. W. Silverman, “Wavelet decomposition approaches to statistical inverse problems,” *Biometrika*, **85**, No. 1, 115–129 (1998).
5. A.V. Markin and O. V. Shestakov, “Asymptotics of the risk estimate for the threshold processing of wavelet-vaguelette coefficients in a tomography problem,” *Inform. Appl.*, **4**, No. 2, 36–45 (2010).
6. O. V. Shestakov, “Mean-square thresholding risk with a random sample size,” *Inform. Appl.*, **12**, No. 3, 14–17 (2018).
7. O. V. Shestakov, “Averaged probability of the error in calculating wavelet coefficients for the random sample size,” *J. Math. Sci.*, **237**, No. 6, 826–830 (2019).
8. S. Mallat, *A Wavelet Tour of Signal Processing*, Academic Press, New York (1999).
9. A. A. Eroshenko and O. V Shestakov, “Asymptotic properties of risk estimate in the problem of reconstructing images with correlated noise by inverting the Radon transform,” *Inform. Appl.*, **8**, No. 4, 32–40 (2014).
10. D. Donoho, I. M. Johnstone, G. Kerkyacharian, and D. Picard, “Wavelet shrinkage: Asymptopia?” *J. Roy. Stat. Soc. Ser. B*, **57**, No. 2, 301–369 (1995).
11. T. Cai and L. Brown, “Wavelet estimation for samples with random uniform design,” *Stat. Prob. Lett.*, **42**, 313–321 (1999).