Mean Square Error Estimation in Thresholding

Soosan Beheshti, *Senior Member, IEEE*, Masoud Hashemi, *Student Member, IEEE*, Ervin Sejdić, *Member, IEEE*, and Tom Chau, *Senior Member, IEEE*

***Abstract*—We present a novel approach to estimating the mean square error (MSE) associated with any given threshold level in both hard and soft thresholding. The estimate is provided by using only the data that is being thresholded. This adaptive approach provides probabilistic confidence bounds on the MSE. The MSE bounds can be used to evaluate the denoising method. Our simulation results confirm that not only does the method provide an accurate estimate of the MSE for any given thresohlding method, but the proposed method can also search and find an optimum threshold for any noisy data with regard to MSE.**

***Index Terms*—Confidence bounds, estimation evaluation, mean square error (MSE), thresholding.**

I. Introduction

ESTIMATING a set of unknown parameters corrupted by additive noise is one of the most important problems in science and engineering. A common measure for the quality of the estimators is the mean-square error (MSE). In many areas (e.g., estimation, denoising, modeling), the aim is to estimate and/or minimize a form of the MSE. When a suboptimal estimation algorithm in the sense of MSE is used, it is of our interest to compare the resultant MSE to the minimum (optimum) MSE (MMSE). However, computing the MMSE is usually cumbersome as well. To overcome this problem various application-dependent methods have been proposed to estimate bounds for MMSE. For example, in [1] a lower bound was provided for error estimation of diffusion filters which is based on covariance inequality and Cramer–Rao bounds [2]. Similar developments can be found in [3]–[6] which are geared towards estimating a desired unknown parameter from an observed pulse-frequency modulated signal. MSE estimators can also be found in applications dealing with linear estimation and regression models (e.g., [7], [8]). Nevertheless, generalizations of these findings are often limited as they often deal with specific applications, classes of signals, and/or a range of signal to noise ratio. In particular, these previous findings are generally not applicable in denoising applications, where estimation of MSE for various threshold values is desired.

Manuscript received October 05, 2010; revised November 18, 2010; accepted November 21, 2010. Date of publication December 10, 2010; date of current version December 23, 2010. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Markku Renfors.

S. Beheshti is with the Department of Electrical and Computer Engineering, Ryerson University, Toronto, ON M5B 2K3 Canada (e-mail: soosan@ee.ryerson.ca).

M. Hashemi is with the Institute of Biomaterials and Biomedical Engineering (IBBME), University of Toronto, Toronto, ON M5S 1A1 Canada (e-mail: sayedmasoud.hashemiamroabadi@utoronto.ca).

E. Sejdić is with the Division of Gerontology, Beth Israel Deaconess Medical Center and Harvard Medical School, Harvard University, Boston, MA 02215 USA (e-mail: esejdic@bidmc.harvard.edu).

T. Chau is with Bloorview Research Institute and the IBBME, University of Toronto, Toronto, ON M5S 1A1 Canada (e-mail: tom.chau@utoronto.ca).

Digital Object Identifier 10.1109/LSP.2010.2097590

In this letter, we propose a method for the MSE estimation in denoising applications. Different denoising techniques exist for removing noise from the desired data, ranging from linear transformations such as Wiener filters [9] to modern approaches that are mostly concentrated on using wavelet coefficients and shrinkage for noise cancelation [10]–[14]. Our development is in line with these modern approaches. Specifically, our method is applicable for any thresholding technique and any type of signal structure. To achieve this goal, we first study the structure of the MSE in a thresholding scenario. Then, we provide probabilistic bounds on the MSE by using only the available noisy data. A similar method has been presented for MSE estimation for subspace selection and linear modeling (e.g., [15] and [16]). To develop an MSE estimator for the purpose of thresholding, the proposed method is based on the same fundamental principles as this previous technique. Here we provide probabilistic confidence bounds on the MSE as a function of the available data.

The letter is arranged as follows. Section II summarizes the background of thresholding and formulates the problem under consideration. In Section III, the structure of MSE in thresholding is provided. Section IV introduces the method for the MSE estimation by using only the available noisy data, while Section V covers the simulation results. Section VI draws the concluding remarks.

II. Problem Statement and Motivation

The unknown noiseless data $\bar{y}$ is corrupted by an additive white Gaussian noise w. The noisy data $y^N = [y(1),\ldots,y(N)]^T$ of length N:

$$y(n)=\bar{y}(n)+w(n) \tag{1}$$

is available for $n=1,2,\ldots,N$. The additive noise $w(n)$ is a sample of a zero mean random variable $W(n)$ with variance σ_w^2. The data is projected onto another orthogonal basis that presents the noiseless data with possibly fewer nonzero coefficients than the data length. The most popular example of such a basis is the class of orthogonal wavelets. The associated coefficients of the data in the new bases are the result of the following inner product

$$\theta(i)=\langle s_i, y^N\rangle=\bar{\theta}(i)+v(i) \tag{2}$$

where s_i, $1\le i\le N$, are the complete orthonormal basis, θ and $\bar{\theta}$ are the coefficients of noisy and noiseless data, and v is the coefficient of the additive noise. Due to the orthonormality of the transformation, the transformed additive noise v is zero mean with the same noise variance σ_w^2.

Thresholding is classified into the two types of hard and soft. Assume that the value of threshold is τ, then we denote the hard thresholded coefficients with $\hat{\theta}_{\tau,H}(i)$ and the soft thresholded version with $\hat{\theta}_{\tau,S}(i)$. We use the notation of $\hat{\theta}_{\tau}$ where both hard and soft thresholding satisfy conditions.

Important Notations: In this letter, the superscript N represent the length of a vector; if X is a random variable, its sample is shown with x.

A. Thresholding and Reconstruction Error

After thresholding, the estimate of noiseless data is given by

$$\hat{y}_\tau^N = \sum_{i=1}^{N} \hat{\theta}_\tau(i) s_i \tag{4}$$

where the s_i is the orthonormal basis. Using the available threshold, the related reconstruction error is

$$z_\tau = \frac{1}{N} \left\| \bar{y}^N - \hat{y}_\tau^N \right\|_2^2 = \frac{1}{N} \left\| \bar{\theta}^N - \hat{\theta}_\tau^N \right\|_2^2. \tag{5}$$

The second equality is due to orthonormality of the transformation basis. The main concern in thresholding is how to choose a proper threshold. One important factor in the thresholding methods has been the mean square reconstruction error (MSE) ([12], [18]):

$$\mathrm{MSE} = \mathrm{E}(Z_\tau) = \frac{1}{N} E\left(\left\| \bar{\theta}^N - \hat{\Theta}_\tau^N \right\|_2^2 \right). \tag{6}$$

Using any method of thresholding, this error can be used for the evaluation purposes. Additionally, the MSE estimates amongst various thresholding approaches enable us to choose an optimal method. In [11], an upper bound for this error is estimated, and an optimal threshold based on analysis of this upper bound is proposed. Here we provide probabilistic upper and lower bounds on the MSE. The objective is to use the observed data and the knowledge about the additive noise to provide bounds on this error.

III. Structure of the MSE in Thresholding

Using either hard or soft thresholding and a threshold value, τ, a corresponding integer m exists such that $N-m$ coefficients are set to zero. To analyze the structure of the reconstruction error for a given threshold τ, we segment the noiseless coefficients as follows $\bar{\theta} = \left[\begin{smallmatrix} \bar{\theta}_\tau \\ \Delta_\tau \end{smallmatrix}\right]$ where Δ_τ is a vector of length $N-m$, corresponding to the coefficients that are set to zero, and $\bar{\theta}_\tau$ is the vector containing the remaining m coefficients of the noise-free data. Consequently, the observed data can be reorganized as follows:

$$y^N = [A_\tau \quad B_\tau] \begin{bmatrix} \bar{\theta}_\tau \\ \Delta_\tau \end{bmatrix} + w^N \tag{7}$$

where the columns of matrix A_τ are s_i for which the associated absolute values of coefficients are larger than τ, and the columns of matrix B_τ are basis vectors for which the absolute values of coefficients are smaller than τ. In the following we provide the structure of the reconstruction error as a function of this segmentation of the data.

A. Hard Thresholding

The coefficients' estimates in hard thresholding, according to (7), are

$$\hat{\theta}_{\tau,H} = \begin{bmatrix} A_\tau^T y^N \\ 0_{(N-m\times 1)} \end{bmatrix} = \begin{bmatrix} \bar{\theta}_\tau + A_\tau^T w^N \\ 0_{(N-m\times 1)} \end{bmatrix}. \tag{8}$$

Therefore, the reconstruction error is

$$z_{\tau,H} = \frac{1}{N} \|\bar{\theta} - \hat{\theta}_{\tau,H}\|_2^2 = \frac{1}{N} \left\| A_\tau^T w^N \right\|_2^2 + \frac{1}{N} \|\Delta_\tau\|_2^2. \tag{9}$$

Note that $A_\tau^T w^N$ are the first m elements of the projected noise, v in (2). Therefore, we have $(1/N)\|A_\tau^T w^N\|_2^2 = (1/N)\sum_{i=1}^{m} v^2(i)$ and since v has a Gaussian distribution, this is a sample of an mth order Chi-square distribution. As a result, the reconstruction error is a sample of a chi-square random variable with the following mean and variance:

$$\mathrm{MSE}_H = \mathrm{E}(Z_{\tau,H}) = \frac{m}{N}\sigma_w^2 + \frac{1}{N}\|\Delta_{S_m}\|_2^2 \tag{10}$$

$$\mathrm{var}(Z_{\tau,H}) = \frac{2m}{N^2} \left(\sigma_w^2\right)^2. \tag{11}$$

B. Soft Thresholding

The coefficients' estimates in soft thresholding, according to (7), is

$$\hat{\theta}_{\tau,S} = \begin{bmatrix} \bar{\theta}_\tau + A_\tau^T w^N - \tau \mathrm{sgn}\left(A_\tau^T y^N\right) \\ 0_{(N-m\times 1)} \end{bmatrix}. \tag{12}$$

Therefore, the reconstruction error is

$$z_{\tau,S} = \frac{1}{N}\|\bar{\theta} - \hat{\theta}_{\tau,S}\|_2^2 \tag{13}$$

$$= \frac{1}{N} \left\| A_\tau^T w^N - \tau \mathrm{sgn}\left(A_\tau^T y^N\right) \right\|_2^2 + \frac{1}{N}\|\Delta_\tau\|_2^2. \tag{14}$$

Since this value is a sample of a noncentral chi-square random variable with the following mean and variance:

$$\mathrm{MSE}_S = \mathrm{E}(Z_{\tau,S}) = \frac{m}{N}\left(\sigma_w^2 + m\tau^2\right) + \frac{1}{N}\|\Delta_\tau\|_2^2 \tag{15}$$

$$\mathrm{var}(Z_{\tau,S}) = \frac{2m}{N^2}\left(\sigma_w^2\right)^2 + \frac{4m}{N^2}\left(\sigma_w^2 \tau^2\right). \tag{16}$$

Details can be found in Appendix A.

IV. Probabilistic Bounds on the Thresholding MSE

In the previous section it was shown that the reconstruction errors for both hard and soft thresholding are samples of mth order Chi-square distribution. To fully define these distributions, it is enough to have their mean and variance. As it is provided, means in (10) and (15), and variances in (11) and (16), are functions of the data length, the threshold value, the additive noise variance, and $(1/N)\|\Delta_\tau\|_2^2$. The threshold value and data length are available. An estimate of the noise variance is needed in many denoising approaches and methods such as median absolute deviation (MAD) are used to estimate this value. The main challenge here is to estimate $(1/N)\|\Delta_\tau\|_2^2$ by *only* using the observed noisy data. To provide an estimate, we adopt a similar approach to what is proposed in [15]. The procedure is as follows. Use the available threshold to calculate the following (hard) coefficient error

$$x_\tau = \frac{1}{N}\left\| \theta^N - \hat{\theta}_{\tau,H}^N \right\|_2^2 = \frac{1}{N} \left\| \begin{bmatrix} 0_{(m\times 1)} \\ \Delta_\tau + B_\tau^T w^N \end{bmatrix} \right\|_2^2 \tag{17}$$

which is the distance between the coefficients of the observed noisy data and coefficients of a hard thresholded version of the same data. This error is a sample of a noncentral Chi-square distribution, X_τ, with the following mean and variance

$$E(X_\tau) = \left(1 - \frac{m}{N}\right)\sigma_w^2 + \frac{1}{N}\|\Delta_\tau\|_2^2 \tag{18}$$

$$\mathrm{var}(X_\tau) = \frac{2}{N}\left(1 - \frac{m}{N}\right)\left(\sigma_w^2\right)^2 + \frac{4\sigma_w^2}{N^2}\|\Delta_\tau\|_2^2. \tag{19}$$

Details of this calculation are analogous to those provided in Appendix A. Using a table of Chi-square distribution we can

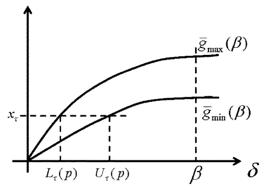

Fig. 1. Finding bounds on $(1/N)\|\Delta_\tau\|_2^2$ by using the available x_τ.

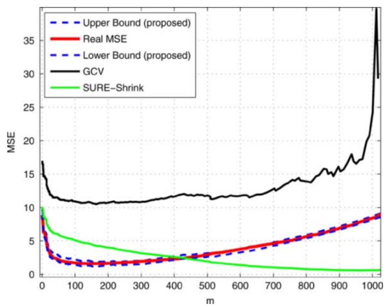

Fig. 2. MSE bounds provided by the proposed method, GCV, and Sure-Shrink for Blocks signal with $\mathrm{SNR} = 0$ dB as a function of m, the number of coefficients that are not set to zero by the thresholding (introduced in Section III).

validate the value of $\|\Delta_\tau\|_2^2$ with any confidence probability p such that the observed sample of data error is around its mean with probability p. This will provide the probabilistic bounds of the form

$$L_\tau(p) \leq \frac{1}{N}\|\Delta_\tau\|_2^2 \leq U_\tau(p). \tag{20}$$

The method is further explained in the following section (Section IV-A). Using these probabilistic bounds, the MSE of hard or soft thresholding, (10) and (15), are bounded as follows

$$\frac{m}{N}\sigma_w^2 + L_\tau(p) \leq \mathrm{MSE}_H \leq \frac{m}{N}\sigma_w^2 + U_\tau(p) \tag{21}$$

$$\frac{m}{N}\left(\sigma_w^2 + m\tau^2\right) + L_\tau(p) \leq \mathrm{MSE}_S \leq \frac{m}{N}\left(\sigma_w^2 + m\tau^2\right) + U_\tau(p). \tag{22}$$

A. Probabilistic Bounds on $(1/N)\|\Delta_\tau\|_2^2$

Here we show how the one available sample of random variable X_τ can be used to find probabilistic bounds on $(1/N)\|\Delta_\tau\|_2^2$.

Fig. 1 shows the procedure for finding the upper and lower bounds. Consider a chi-square random variable G_δ with mean and variance of the following form

$$E(G_\delta) = \left(1 - \frac{m}{N}\right)\sigma_w^2 + \delta \tag{23}$$

$$\mathrm{var}(G_\delta) = \frac{2}{N}\left(1 - \frac{m}{N}\right)\left(\sigma_w^2\right)^2 + \frac{4\sigma_w^2}{N}\delta. \tag{24}$$

The graph shows the behavior of confidence bounds on this distribution with probability p as the value of δ changes. For example if δ takes a value such as β, shown in the graph, the probabilistic bounds on the observed data are shown to be $\bar{g}_{\min}(\beta)$ and $\bar{g}_{\max}(\beta)$. We validate the set of δs that include the observed x_τ in their confidence region

$$L_\tau(p) = \min_\delta \bar{g}_{\min}(\delta) \leq x_\tau \leq \bar{g}_{\max}(\delta) \tag{25}$$

TABLE I
FOR EACH THRESHOLDING METHOD, EACH SNR, AND EACH SIGNAL, THE TOP NUMBER IS THE LOWER BOUND AND THE MIDDLE NUMBER IS THE MSE AND THE BOTTOM NUMBER IS THE UPPER BOUND (ESTIMATED MSE)

	Universal Threshold (Hard)		
	SNR		
	0	5	10
Blocks	2.5203 2.5296 2.5348	0.6062 0.6102 0.6579	0.2050 0.2114 0.2231
Bumps	0.5210 0.6032 0.6909	0.2332 0.2715 0.3393	0.1322 0.1756 0.2141
QuadChirp	0.3870 0.4517 0.5347	0.3907 0.4507 0.5182	0.3998 0.4490 0.4728
MishMash	1.1590 1.3683 1.5694	1.1823 1.3420 1.5352	1.2803 1.3417 1.3938
Swallow-like	0.0322 0.0487 0.0516	0.0268 0.0288 0.0295	0.0179 0.0183 0.0184

	VisuShrink (Soft)			SureShrink (Soft)		
	SNR			SNR		
	0	5	10	0	5	10
Blocks	1.6008 1.7425 1.8282	0.9924 1.1198 1.2595	0.6993 0.7334 0.7602	7.4453 7.5544 7.6274	1.5329 1.5737 1.5964	0.0591 0.1116 0.2117
Bumps	0.5778 0.7828 0.9676	0.2387 0.5517 0.6638	0.3243 0.3955 0.4425	1.8922 2.0306 2.0584	0.0965 0.1250 0.1903	0.0393 0.0732 0.0987
QuadChirp	0.3876 0.4680 0.5471	0.4012 0.4695 0.5222	0.4651 0.4704 0.4886	0.2984 0.3714 0.4253	0.3354 0.3532 0.3584	0.3491 0.3550 0.3562
MishMash	1.3076 1.4196 1.5475	1.2866 1.4226 1.5131	1.4035 1.4253 1.4393	0.8595 0.8806 0.9284	0.6472 0.6637 0.6958	0.7931 0.8848 0.9384
Swallow-like	0.0661 0.0731 0.0755	0.0552 0.0581 0.0592	0.0416 0.0426 0.0434	0.0232 0.0233 0.0238	0.0109 0.0123 0.0131	0.0102 0.0105 0.0116

$$U_\tau(p) = \max_\delta \bar{g}_{\min}(\delta) \leq x_\tau \leq \bar{g}_{\max}(\delta). \tag{26}$$

V. SIMULATION RESULTS

We provide examples of MSE estimation for both hard and soft thresholding methods. The test signals are prototype signals named Blocks, Bumps, QuadChirp, and Mishmash that are introduced in [12] and cover a range of sparse and non-sparse signals. In addition, another test signal is a swallow-like signal which is much more detailed than the other test signals [17]. The swallow-like signal is used in primary evaluations of pioneer signal processing techniques geared towards the analysis of dysphagia (swallowing difficulties). For the wavelet transform, five level decomposition with Haar wavelet is chosen for this experiment. For our MSE estimation, we set the confidence probability to be $p = 0.9$. Fig. 2 shows how the MSE bounds are provided by the method for different values of threshold for Blocks signal at $\mathrm{SNR} = 0$ dB. It also shows the MSE estimate by SURE-Shrink [12] and Generalized Cross Validation (GCV) [13]. As the figure shows, the bounds are tight even for such low SNR. The performance of the method continues to improve as the SNR grows. It is important to note that SURE-Shrink performs poorly for SNRs even up to 10 dB for all the signals and shows a gradual improvement for some of the tested signals as SNR grows above 10 dB.

Table I shows the results of MSE estimation for hard and soft thresholding approaches. The MSE estimates are the probabilistic worse case estimates, in the form of the provided upper bound in (21) and (22). Each result in the table is the average of 100 trials. As the table illustrates, the MSE estimates are comparable with the true MSEs for each of the thresholding methods. For example, when universal thresholding is used for Blocks at $\mathrm{SNR} = 0$ dB, the true MSE, that is calculated by using the noiseless data, is 2.5296, whereas the MSE estimate, without

TABLE II
FOR EACH SNR AND EACH SIGNAL, THE TOP NUMBER IS MSE AND THE BOTTOM NUMBER IS THE ESTIMATED MSE

	MSEEM(Soft Threshold)			MSEEM (Hard Threshold)		
Signal	SNR			SNR		
	0	5	10	0	5	10
Blocks	1.3383 1.3641	0.6717 0.6812	0.1106 0.1807	1.0986 1.3008	0.6094 0.6637	0.1295 0.1929
Bumps	0.5509 0.5706	0.0933 0.0939	0.0729 0.0895	0.4875 0.5120	0.2141 0.2428	0.1083 0.1277
QuadChirp	0.2551 0.2553	0.1021 0.1035	0.0411 0.0412	0.3624 0.3871	0.1440 0.1601	0.0490 0.0515
MishMash	0.8128 0.8177	0.3720 0.3730	0.1473 0.1474	1.3021 1.3413	0.4573 0.6003	0.1596 0.2053
Swallow-like	0.0159 0.0160	0.0065 0.0065	0.0027 0.0027	0.0143 0.0158	0.0059 0.0068	0.0024 0.0027

using the noiseless data, is 2.5348. As the table confirms, the proposed method can be used for comparison and evaluation of different thresholding methods for any available noisy data. Consequently, we can use a range of thresholds and estimate the MSE associated to those thresholds. Comparison of these MSE estimates provides the optimum threshold regarding the MSE. We denote this adaptive thresholding method by MSE EstiMator (MSEEM) thresholding method. Table II shows the results of using MSEEM thresholding for the same data used in the previous table. Not only is the MSE estimate still very close to the exact MSE, but as it was expected, compared to Table I, the MSEEM method also provides thresholds that minimize the MSE further than the existing thresholding methods in both soft and hard thresholding. It is noteworthy to mention that while the proposed method outperforms the existing approaches, the complexity of the algorithm is comparable to that of the existing approaches.

VI. CONCLUSION

In this letter, we derived probabilistic bounds of the MSE in denoising using hard and soft thresholding. The bounds can be utilized to evaluate the denoising performance with regards to the MSE. The adaptive bounds are functions of only the observed noisy data itself. For any range of possible thresholds, the method can estimate the MSE. Therefore, it can provide the optimum threshold that minimizes the MSE estimate for any available noisy data.

APPENDIX A
RECONSTRUCTION ERROR IN SOFT THRESHOLDING

Since $A_\tau^T w^N$, v in (2), includes the first m elements of the projected noise, the reconstruction error in (14) has the following form

$$z_{\tau,S} = \frac{1}{N}\sum_{i=1}^{m}(v(i)+e(i))^2 + \frac{1}{N}\|\Delta_\tau\|_2^2 \tag{27}$$

where $e(i)$ is such that

$$\sum_{i=1}^{m} e^2(i) = \left\|\tau \operatorname{sgn}\left(A_\tau^T y^N\right)\right\|_2^2 = m\tau^2. \tag{28}$$

Therefore, a sum of form $\sum_{i=1}^{m}(v(i)+e(i))^2$ is a noncentral chi-square of order m with mean and variance

$$\mathrm{E}\left(\sum_{i=1}^{m}(v(i)+e(i))^2\right) = m\sigma_w^2 + \sum_{i=1}^{m} e^2(i) \tag{29}$$

$$\mathrm{var}\left(\sum_{i=1}^{m}(v(i)+e(i))^2\right) = 2m(\sigma_w^2)^2 + 4\sigma_w^2\sum_{i=1}^{m} e^2(i). \tag{30}$$

Consequently, the expected value and variance of the reconstruction error are (15) and (16).

REFERENCES

[1] B. Z. Bobrovski and M. Zakai, "A lower bound on the estimation error for certain diffusion problems," *IEEE Trans. Inf. Theory*, vol. IT-22, no. 1, pp. 45–52, Jan. 1976.

[2] A. Renaux, P. Forster, P. Larzabal, C. D. Richmond, and A. Nehorai, "A fresh look at the bayesian bounds of the Weiss–Weinstein family," *IEEE Trans. Signal Process.*, vol. 56, no. 11, pp. 5334–5352, Nov. 2008.

[3] S. Bellini and G. Tartara, "Bounds on error in signal parameter estimation," *IEEE Trans. Commun.*, vol. COM-22, no. 3, pp. 340–342, Mar. 1974.

[4] D. Chazan, M. Zakai, and J. Ziv, "Improved lower bounds on signal parameter estimation," *IEEE Trans. Inf. Theory*, vol. 21, no. 1, pp. 90–93, 1975.

[5] K. L. Bell, Y. Steinberg, Y. Ephraim, and H. L. Van Trees, "Extended Ziv–Zakai lower bound for vector parameter estimation," *IEEE Trans. Inf. Theory*, vol. 43, no. 2, pp. 624–637, Jan. 1997.

[6] J. Ziv and M. Zakai, "Some lower bounds on signal parameter estimation," *IEEE Trans. Inf. Theory*, vol. 15, no. 3, pp. 386–391, May 1969.

[7] K. Kim, "MSE estimation of homogeneous linear estimator in stratified multi-stage sampling," in *Proc. Annu. Meeting of the Amer. Statist. Assoc.*, Aug. 2001.

[8] M. Verbeek and T. Nijman, "Minimum MSE estimation of a regression model with fixed effects from a series of cross-sections," *J. Econ.*, vol. 59, pp. 125–136, Sep. 1993.

[9] J. Chen, J. Benesty, Y. Huang, and S. Doclo, "New insights into the noise reduction Wiener filter," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 14, no. 4, pp. 1218–1234, Jul. 2006.

[10] S. G. Chang, B. Yu, and M. Vetterli, "Adaptive wavelet thresholding for image denoising and compression," *IEEE Trans. Image Process.*, vol. 9, no. 9, pp. 1532–1546, Sep. 2000.

[11] D. L. Donoho and I. M. Johnstone, "Ideal spatial adaption via wavelet shrinkage," *Biometrika*, vol. 81, no. 3, pp. 425–455, 1994.

[12] D. L. Donoho and I. M. Johnstone, "Adapting to unknown smoothness via wavelet shrinkage," *J. Amer. Statist. Assoc.*, vol. 90, no. 432, pp. 1200–1224, Dec. 1995.

[13] M. Jansen, M. Malfait, and A. Bultheel, "Generalized cross validation for wavelet thresholding," *Signal Process.*, vol. 56, no. 1, pp. 33–44, Jan. 1997.

[14] G. P. Nason, "Wavelet shrinkage using cross validation," *J. Roy. Statist. Soc.*, ser. Series B, vol. 58, pp. 463–479, 1996.

[15] S. Beheshti and M. A. Dahleh, "A new information-theoretic approach to signal denoising and best basis selection," *IEEE Trans. Signal Process.*, vol. 53, no. 10, pp. 3613–3624, Oct. 2005.

[16] S. Beheshti and M. A. Dahleh, "Noisy data and impulse response estimation," *IEEE Trans. Signal Process.*, vol. 58, no. 2, pp. 510–521, Feb. 2010.

[17] E. Sejdić, C. M. Steele, and T. Chau, "A procedure for denoising of dual-axis swallowing accelerometry signals," *Physiol. Meas.*, vol. 31, no. 1, pp. N1–N9, Jan. 2010.

[18] H. Krim, D. Tucker, S. Mallat, and D. Donoho, "On denoising and best signal representation," *IEEE Trans. Inf. Theory*, vol. 45, no. 7, pp. 2225–2238, Nov. 1999.