A Step-Down Test Procedure for Wavelet Shrinkage Using Bootstrapping

MUNWON LIM[1], OLUFEMI A. OMITAOMU[2], (Senior Member, IEEE), AND SUK JOO BAE[1], (Member, IEEE)

[1]Department of Industrial Engineering, Hanyang University, Seoul 04763, South Korea
[2]Oak Ridge National Laboratory, Computational Sciences and Engineering Division, Oak Ridge, TN 37831, USA

Corresponding author: Suk Joo Bae (sjbae@hanyang.ac.kr)

ABSTRACT Wavelet thresholding (or shrinkage) attempts to remove the noises existing in the signals while preserving inherent pattern characteristics in the reconstruction of true signals. For data-denoising purpose, we present a new wavelet thresholding procedure which employs the step-down testing idea of identifying active contrasts in unreplicated fractional factorial experiments. The proposed method employs bootstrapping methods to a step-down test for thresholding wavelet coefficients. By introducing the concept of a false discovery error rate in testing wavelet coefficients, we shrink the wavelet coefficients with p-values higher than the error rate. The error rate controls the expected proportion of wrongly accepted coefficients among chosen wavelet coefficients. Bootstrap samples are used to approximate the p-value for computational efficiency. We also present some guidelines for selecting the values of hyper-parameters which affect the performance in the step-down thresholding procedure. Based on some common testing signals and an air-conditioner sounds example, the comparison of our proposed procedure with other thresholding methods in the literature is performed. The analytical results show that the proposed procedure has a potential in data-denoising and data-reduction in a variety of signal reconstruction applications.

INDEX TERMS Bootstrap aggregating, complex wavelet transform, data-denoising, step-down test, wavelet shrinkage.

I. INTRODUCTION

Wavelet transform (WT), due to its excellent localization property, has rapidly grown up as a powerful tool for analyzing noisy signals that arise in various fields of applications, e.g., denoising, function estimation, change-point detection, inverse problem, principal components, and discrimination analysis [23]. One reason for this popularity is the ''sparseness'' property of the WT, that is, signals from a wide range of applications can have a parsimonious representation through only a few significant coefficients in wavelet series [3]. The WT resembles the fast Fourier transform (FFT). However, the WT can successfully model irregular data patterns better than other statistical methods such as spline and polynomial regression, as well as the FFT [12], [14]. In addition, because the WT provides an excellent multi-resolution approximation

through dilation and translation, it can effectively extract time-frequency features of noise signals [17].

For signal denoising purpose, wavelet *thresholding* (or *shrinkage*) attempts to remove the noises existing in the signals while preserving inherent signal characteristics, regardless of its frequency content. To date, a number of wavelet thresholding methods have been developed to apply such good properties of the WTs to data-denoising or data reduction. From a statistical point of view, the wavelet thresholding closely relates to multiple hypotheses tests where each coefficient is tested whether it is statistically significant or not [1], [9]. Due to the sparse property of wavelets, it is reasonable to assume that only a few wavelet coefficients contain inherent information about the real signal, while other coefficients appear as a consequence of signal's corruption by random noises. Accordingly, the wavelet coefficients that are ''significantly different from zero'' can be used in the reconstruction to original signals. The purpose of wavelet

The associate editor coordinating the review of this manuscript and approving it for publication was Zhaojun Li.

thresholding is to extract these significant coefficients by ignoring others through a threshold value. Because the results of a hypothesis test should guide the choice of significant coefficients, a stronger control for noise levels is required lest truly insignificant coefficients should be used in the model building [2]. Many thresholding methods in the literature shrink insignificant coefficients by directly calculating a threshold value from an estimate of the noise variance, or assuming that the variance is known *a priori*. Considering a white-noise process for noise errors, for example, Donoho and Johnstone [9] proposed a universal thresholding method to shrink the coefficients resulted from the noises. However, if the estimate of error variance is highly biased or distributional assumptions for the errors are mis-specified, the threshold value for wavelet shrinkage will be misguided, thus at risk of shrinking significant wavelet coefficients as insignificant and vice versa.

In this paper, we propose a statistical method for thresholding the (complex-valued) wavelet coefficients based on a step-down testing procedure using bootstrapping. The proposed method employs Venter and Steel's idea [22] on identifying active contrasts from unreplicated fractional experiments. The experimental design approach for wavelet thresholding has many features that make it attractive from a practical point of view. The contribution or this work can be summarized as follows:

- The proposed method has an advantage over other wavelet thresholding methods in that the method indirectly deals with error variance by introducing scale-invariant ordinal statistics so that the undistorted shrinkage is provided.
- The proposed method is capable of considering a false discovery error rate which makes the wavelet shrinkage less stringent in multiple hypothesis testing problems.
- The amount of noise removal is adjustable by introducing hyper-parameters that can be controlled according to user-specification, where the user can modulate the degree of data reduction.

In brief, our proposed method is simple and easily interpretable since it involves only two quantities; the assumed least number of insignificant coefficients and a significance level of testing coefficients. Additionally, the approach provides a flexible guidance for checking a ''sparseness'' of wavelet coefficients by computing *p*-values, then by comparing them with the significance level. Coping with the need for aggressively reducing the dimension for a high-dimensional data set, as in an air-conditioner example in Section IV-B, the suggested procedure facilitates the control of shrinkage ratios via a user-specified error rate.

The paper is organized as follows. In Section II, we briefly review complex wavelet transforms and several wavelet thresholding methods. Section III proposes a wavelet thresholding based on step-down testing procedure. The *p*-values of the hypotheses tests are calculated using bootstrap-based approximation method. In Section IV, we provide analytical results to illustrate our procedure and compare its

performance with those of other existing wavelet thresholding methods via a variety of simulations and a real example of air-conditioner noise sounds. Some concluding remarks are presented in Section V with the discussion about future research directions.

II. COMPLEX WAVELET TRANSFORMS AND THRESHOLDING METHODS
A. COMPLEX WAVELET TRANSFORMS
Wavelets are basis functions that allow the transformation of signals from their original domain to another domain in which some operations can be performed in a flexible manner. Such transformation provides opportunities for effective data-denoising and data-reduction processes. In a variety of signal processing applications, discrete wavelet transforms (DWTs) have been widely used due to the easiness of its fast, local, sparse, and decorrelated multi-resolution analysis of signals [11]. However, DWTs may suffer from serious applicative problems such as shift sensitivity arising from downsamplers in the implementation [21] and no phase information [13].

As an alternative, complex wavelet transforms (CWTs) extend the idea of real-valued wavelet transform by employing complex-valued filters in place of real-valued filters. In particular, Daubechies complex wavelet transform has been well known as an effective decomposition method for capturing the boundary or abrupt changes of signals or images. It can be controlled to be symmetric and shift-invariant to reduce the risk of a distortion in the wavelet domains and to easily handle the boundary problems of the object [8].

Suppose that the translation set of dilated scaling and wavelet functions

$$\phi_{L,k}(t) = 2^{L/2}\phi(2^L t - k) \quad \text{and} \tag{1}$$
$$\psi_{j,k}(t) = 2^{j/2}\psi(2^j t - k), \tag{2}$$

are orthonormal bases at resolution level j and location k in $L^2(\mathbb{R})$, where L denotes the lowest decomposition level and $\mathbb{Z}$ is the set of all integers. Any function $\tilde{f}(t)$ in $L^2(\mathbb{R})$ can be represented as

$$\tilde{f}(t) = \sum_{k=0}^{2^L-1} c_{L,k}\phi_{L,k}(t) + \sum_{j=L}^{J}\sum_{k=0}^{2^j-1} d_{j,k}\psi_{j,k}(t), \tag{3}$$

for $J \geq L$, where the coefficients $c_{L,k} = \int_{\mathbb{R}} f(t)\phi_{L,k}(t)\,\mathrm{d}t$ are considered to be the coarsest-level coefficients characterizing smoother data patterns and $d_{j,k} = \int_{\mathbb{R}} f(t)\psi_{j,k}(t)\,\mathrm{d}t$ are viewed as the finer-level coefficients which capture high frequency oscillations of data patterns [19]. The two wavelet coefficients $c_{L,k}$ and $d_{j,k}$ also consist of both real and imaginary parts as

$$c_{L,k} = \Re(c_{L,k}) + i \cdot \Im(c_{L,k}) \quad \text{and} \tag{4}$$
$$d_{j,k} = \Re(d_{j,k}) + i \cdot \Im(d_{j,k}), \tag{5}$$

respectively.

The simplest version of wavelet thresholding is typically expressed as the estimation of a signal ($f(t_i)$) given a noise data ($y(t_i)$) as

$$y(t_i) = f(t_i) + \epsilon_i, \quad i = 1, \ldots, N, \quad (6)$$

where the errors ϵ_i are assumed independently $\mathcal{N}(0, \sigma^2)$ distributed. We can rewrite Eq. (6) as $\mathbf{y} = \mathbf{f} + \boldsymbol{\epsilon}$, for $\mathbf{f} \equiv (f(t_1), \ldots, f(t_N))^T$, $\mathbf{y} \equiv (y(t_1), \ldots, y(t_N))^T$, and $\boldsymbol{\epsilon} \equiv (\epsilon_1, \ldots, \epsilon_N)^T$. Then, the CWT of $\mathbf{y}$ is defined as

$$\mathbf{d} = \mathbf{Wy} = \mathbf{Wf} + \mathbf{W}\boldsymbol{\epsilon}, \quad (7)$$

where $\mathbf{W}$ is the orthonormal ($N \times N$) matrix corresponding to the CWT. Here, $\mathbf{d}$ is the combined set of ($N \times 1$) wavelet coefficients consisting of real and imaginary parts, describing the features of the original function. $\mathbf{d}$ can be represented as: $\mathbf{d} \equiv (\mathbf{c}_L, \mathbf{d}_L, \mathbf{d}_{L+1}, \ldots, \mathbf{d}_J)^T$, through the wavelet coefficients at various subbands (scales), $\mathbf{c}_L \equiv (c_{L,0}, \ldots, c_{L,2^L-1})^T$; $\mathbf{d}_L \equiv (d_{L,0}, \ldots, d_{L,2^L-1})^T$; $\mathbf{d}_{L+1} \equiv (d_{L+1,0}, \ldots, d_{L+1,2^{L+1}-1})^T$; $\cdots$, $\mathbf{d}_J \equiv (d_{J,0}, \ldots, d_{J,2^J-1})^T$. Note that the total number of wavelet coefficients is equal to the number of measurements, that is, $N = 2^{J+1}$. Hereafter, the wavelet coefficients are represented by $\mathbf{d} \equiv (d_1, d_2, \ldots, d_N)^T$ for the simplicity of notation. Using the inverse CWT, the ($N \times 1$) vector of the original signal is reconstructed via

$$\mathbf{y} = \mathbf{W}^{-1}\mathbf{d}. \quad (8)$$

In the CWT, individual components of $\mathbf{W}\boldsymbol{\epsilon} (\equiv \boldsymbol{\epsilon}')$, which are considered as complex-valued random variables, are uncorrelated. However, the real and imaginary parts of $\boldsymbol{\epsilon}'$ are real-valued normal random variables and they can be strongly correlated. Barber and Nason [4] showed that for a unitary matrix $\mathbf{W}$,

$$\text{Cov}\{\mathfrak{Re}(\boldsymbol{\epsilon}'), \mathfrak{Im}(\boldsymbol{\epsilon}')\} = -\sigma^2 \mathfrak{Im}(\mathbf{WW}^T)/2, \quad (9)$$

$$\text{Cov}\{\mathfrak{Re}(\boldsymbol{\epsilon}'), \mathfrak{Re}(\boldsymbol{\epsilon}')\} = \sigma^2\{\mathbf{I}_N + \mathfrak{Re}(\mathbf{WW}^T)\}/2, \quad (10)$$

$$\text{Cov}\{\mathfrak{Im}(\boldsymbol{\epsilon}'), \mathfrak{Im}(\boldsymbol{\epsilon}')\} = \sigma^2\{\mathbf{I}_N - \mathfrak{Re}(\mathbf{WW}^T)\}/2, \quad (11)$$

since $\boldsymbol{\epsilon}$ are $\mathcal{MVN}(0, \sigma^2\mathbf{I}_N)$ distributed. Here $\mathcal{MVN}(\cdot)$ denotes a multivariate normal. The noise error variance σ^2 is generally estimated by the squared mean deviation of the coefficients at the finest resolution level, but Barber and Nason [4] estimated σ^2 by the sum of the squared median absolute deviation of the real and imaginary parts of the finest-leveled coefficients. Finally, $\mathbf{d}$ becomes a bivariate normal vector with mean, $\mathbf{Wf} (\equiv \boldsymbol{\theta})$, and the variance-covariance matrix of Σ which are equal to the elements of the variance-covariance matrix of $\boldsymbol{\epsilon}'$.

B. WAVELET THRESHOLDING METHODS

When the signals of interest are transformed by the wavelets, the meaningful energy is prone to concentrating on the small number of wavelet coefficients which may contain information of real signals. Wavelet thresholding is a denoising method by comparing transformed coefficients with a threshold value, then only the coefficients larger than the threshold are used to reconstruct the denoised signals. Most popular

three thresholding procedures in CWTs are VisuShrink [9], RiskShrink [9], and complex multiwavelet style shrinkage (CMWS) [4].

The VisuShrink threshold method, usually called a "universal threshold", requires an estimation of error variance of the noises to calculate the threshold value. The estimators obtained by reconstructing the function from the remaining coefficients via the threshold value have the risk close to the minimal risk corresponding to the optimal thresholding rule. Sardy [20] developed a universal thresholding method for the CWTs to preserve the additional phase information. Therein, along with the threshold value of $\lambda_{VS} \equiv \hat{\sigma}\sqrt{2\log(N \log N)}$, the hard and soft thresholding rules are given by

$$Visu_h(d_{j,k}^*) = d_{j,k}^* \cdot \mathbb{I}(|d_{j,k}^*| > \lambda_{VS}) \text{ and} \quad (12)$$

$$\frac{d_{j,k}^*}{|d_{j,k}^*|} \cdot \max(|d_{j,k}^*| - \lambda_{VS}, 0), \quad (13)$$

respectively, where $d_{j,k}^*$ is the coefficient at resolution level j and location k. Here, $\mathbb{I}(\cdot)$ denotes the indicator function. According to the rules, the coefficients smaller than the threshold are removed as noises. As a minimax threshold method, RiskShrink minimizes a theoretical upper bound on the asymptotic risk. The minimax method is efficient in picking up abrupt jumps, but at the expense of smoothness. Unlike previous two methods that directly compare the wavelet coefficients with the threshold value, CMWS [4] considers not only the wavelet coefficients, but also the covariance of resolution levels. Based on the complex multi-wavelet shrinkage, this method shrinks the noisy data by calculating the level-weighted statistics so that the information of each resolution level is preserved. Its threshold corresponds to $\lambda_{CMWS} \equiv 2\log N$. The hard and soft thresholding rules for CMWS are

$$CMWS_h(\mathbf{d}_{j,k}^*) = \mathbf{d}_{j,k}^* \cdot \mathbb{I}(\vartheta_{j,k} > \lambda_{CMWS}) \text{ and} \quad (14)$$

$$CMWS_s(\mathbf{d}_{j,k}^*) = \frac{\mathbf{d}_{j,k}^*}{|\mathbf{d}_{j,k}^*|} \cdot \max(\vartheta_{j,k} - \lambda_{CMWS}, 0), \quad (15)$$

for shrinkage statistics $\vartheta_{j,k} = \mathbf{d}_{j,k}^{*T} \Sigma_j^{-1} \mathbf{d}_{j,k}^*$, where $\mathbf{d}_{j,k}^* \sim \mathcal{MVN}_2(\mathbf{d}_{j,k}, \Sigma_j)$. Here, Σ_j denotes the variance-covariance matrix at decomposition level j. The thresholding rules are based on the fact that $\vartheta_{j,k}$ has a non-central χ_2^2 distribution with non-centrality parameter $\mathbf{d}_{j,k}^T \Sigma_j^{-1} \mathbf{d}_{j,k}$. Note that $\vartheta_{j,k}$ follows a central χ_2^2 distribution if $\mathbf{d}_{j,k} = \mathbf{0}$. The CMWS method consistently provides more accurate estimates than the shrinkage methods based on real-valued wavelets. In addition, other data-driven thresholding for the CWTs include Bayesian approaches [4] for determining wavelet thresholding levels, SURE thresholding rule [6], which is based on minimizing Stein's unbiased risk estimate, and the generalized SURE thresholding rule based on the variance estimator [15].

Overall, existing methods shrink the data dimension without any direct control of the degrees of noise removal. However, if we seek to shrink data dimensionality while managing

the noise level of original signals efficiently, it will be better for users to specify a denoising rate by controlling resolution levels of wavelet-domain signals. In the following section, we propose a wavelet step-down thresholding that designates a user-specified error rate by pre-specifying a lower bound on the total number of insignificant coefficients. Moreover, the variance-related statistics are indirectly dealt with so that the de-noising results are robust regardless of the error variance.

III. THE STEP-DOWN THRESHOLDING PROCEDURE

We rewrite the CWT model in (7) to illustrate a wavelet step-down thresholding (WSDT) procedure as

$$\mathbf{d} = \boldsymbol{\theta} + \boldsymbol{\epsilon}'. \tag{16}$$

Due to the orthogonality of $\mathbf{W}$, $\boldsymbol{\epsilon}'$ have the identical structure with $\boldsymbol{\epsilon}$ as iid $\mathcal{MVN}(\mathbf{0}, \boldsymbol{\Sigma})$. Thus, $\mathbf{d} \sim \mathcal{MVN}(\boldsymbol{\theta}, \boldsymbol{\Sigma})$. One can obtain the signal f from the inverse transformation of $\boldsymbol{\theta}$. However, true values of $\boldsymbol{\theta} \equiv (\theta_1, \ldots, \theta_N)$ and σ^2 are unknown and must be estimated from the wavelet coefficients $\mathbf{d}$ only. At this case, the estimation of $\mathbf{d}$ independent of σ^2 is not available. Different estimation methods for σ^2 will lead to distinct shrinkage schemes of wavelet coefficients, resulting in different dimensions of data reduction. In general, small-valued coefficients result from noise data, thus thresholding out these coefficients has an effect of "removing noises". Relatively a few large-valued coefficients can effectively contribute to the reconstruction of true signals. In using any types of wavelet thresholding methods, a main issue is how to choose the threshold value for all of the wavelet coefficients.

The thresholding rule closely associates with identifying active (that is, "non-zero") contrasts in a single replicate of experimental design problems. Consider a linear model for an experimental design, $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, where $\mathbf{Y}$ is a $(N \times 1)$ vector of responses, design matrix $\mathbf{X}$ is orthogonal, and $\boldsymbol{\epsilon}$ is a $(N \times 1)$ vector of random error and assumed to be iid $\mathcal{MVN}(\mathbf{0}, \sigma^2 \mathbf{I}_N)$. In a design of experiment, an ordinary least squares estimator $\hat{\boldsymbol{\beta}} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{Y}$ becomes $\hat{\boldsymbol{\beta}} = \mathbf{X}^T\mathbf{Y}$, which is an orthogonal transformation of the observations. The stepwise elimination of inactive contrasts in the model corresponds to removing components with the smallest absolute value of t-statistics for $\hat{\boldsymbol{\beta}}$. Since the design is orthogonal, the values of remaining $\hat{\boldsymbol{\beta}}$ do not change in the process of elimination. This stepwise deletion procedure characterizing a multiple hypotheses test is considered as a hard-thresholding rule in the case of orthogonal wavelet transforms [24, p. 177]; that is, $\hat{\beta}_i^* = \max(|\hat{\beta}_i| - \lambda, 0)$, where $\hat{\beta}^*$ is an active contrast for the threshold $\lambda(> 0)$.

Introducing the multiple hypotheses-testing approach to wavelet thresholding problems, we seek to build a test consisting of the overall null hypothesis that all the d_i's are zero (referred to as H_0). If H_0 is rejected, the non-zero d_i's that cause the rejection of the hypothesis are identified. The procedure is designed to control the probability of wrongly declaring at least one of insignificant coefficients as significant

using their p-values. Because the presence of falsely declared significant coefficients can cause biased results, it is desirable to control the error rate. The WSDT proceeds downward from the wavelet coefficient with the largest variance-weighted statistic. Through the iteration of bootstrapping, a set of maximum data points satisfying the hypotheses test is constructed. The WSDT procedure involves the following steps:

- Step 1: Define the variance-weighted statistics $\theta_{j,k} = \mathbf{d}_{j,k}^{*T} \Sigma_j^{-1} \mathbf{d}_{j,k}^*$, which are the shrinkage statistics in CMWS method [4], then sort $\theta_{j,k}$'s so that $\{\theta_{(1)} < \theta_{(2)} < \cdots < \theta_{(N)}\}$ are the order statistics of $\boldsymbol{\theta}$.

- Step 2: Define a scale-invariant ratio T_q as

$$T_q = \frac{\theta_{(q)}}{\left(\sum_{i=1}^{l-1} \theta_{(i)}^2 / (l-1)\right)^{1/2}}, \tag{17}$$

where l is a specified lower bound to determine the number of zero-valued wavelet coefficients. Here, q is an index of non-zero coefficients, i.e., $q = l+1, \ldots, N$. The denominator is a fixed scale-equivalent function that depends only on the pre-determined value of l. The scale-invariant ratio is regarded as a random variable of the standardization form that guarantees the evaluation of the coefficients on the same scale [18, p.90]. Section III-B will discuss the guidelines to determine lower bound l and investigate its effect on the overall performance of the WSDT procedure.

- Step 3: Based on the tentative $(N - l)$ non-zero samples, bootstrap T_q's satisfying the following condition

$$\tilde{T}_q^{(b)} \equiv \underset{T_q}{\text{Max}}(T_q \geq T_q^{(b)} \mid H_q), \tag{18}$$

for $q = l+1, \ldots, N$, $b = 1, \ldots, B$, where H_q denotes the parameter configuration in which exactly l number of the wavelet coefficients are zeros. The collection of sampled data, $\tilde{\mathbf{T}}_q \equiv \{\tilde{T}_q^{(1)}, \ldots, \tilde{T}_q^{(B)}\}$, is constructed from the bootstrapping procedure. The distribution of the scale-invariant ratios under H_q does not depend on the variance and we will take the unit variance for convenience when calculating the distribution of $\tilde{T}_q$ [22].

- Step 4: A test statistic for determining the number of insignificant coefficients is achieved using the concept of p-value as

$$\Pr(\tilde{T}_q^{(b)} \geq \lambda_\kappa \mid H_q) \leq \kappa, \quad b = 1, \ldots, B, \tag{19}$$

where κ is the false discovery error rate that controls the expected proportion of wrongly accepted coefficients among chosen wavelet coefficients [5]. The values of the threshold λ_κ are controlled by the false discovery error rate κ. It implies that the level of shrinkage can be controlled by specifying the error rate. If the aggressive thresholding is required, a user can set smaller values of κ. The calculation of the p-values using the bootstrap procedure may be computationally intensive, thus we introduce approximate methods, which will be illustrated in Section III-A.

- Step 5: Extract significant coefficients using either hard or soft thresholding criterion. The soft thresholding criterion is defined by

$$\mathbf{d}_{j,k}^* = \begin{cases} \mathbf{d}_{j,k}^*/|\mathbf{d}_{j,k}^*|(\theta_{j,k} - \lambda_\kappa), & \text{if } \theta_{j,k} > \lambda_\kappa, \\ \mathbf{0}, & \text{if } \theta_{j,k} \le \lambda_\kappa. \end{cases} \quad (20)$$

Alternatively, the hard thresholding criterion is

$$\mathbf{d}_{j,k}^* = \begin{cases} \mathbf{d}_{j,k}^*, & \text{if } \theta_{j,k} > \lambda_\kappa, \\ \mathbf{0}, & \text{if } \theta_{j,k} \le \lambda_\kappa. \end{cases} \quad (21)$$

Hereafter, the procedure based on the soft thresholding rule is called "WSDT$_s$" while the hard thresholding rule is called "WSDT$_h$". In reconstructing the wavelet coefficients, Donoho and Johnstone [9] suggested that the coefficients at the coarse levels should always be included even if these coefficients do not pass the thresholding rule. More flexibility according to applicative problems will be given while adopting their suggestion in this work.

A. APPROXIMATION METHOD FOR COMPUTING p-VALUE
The bootstrap-based approach provides a simple but effective calculation of p-value alternatively. To describe the approximation to compute p-value, let T_q^* be a random variable whose distribution is the same as that of the bootstrapped samples $T_q^{(b)}$ for $b = 1, \ldots, B$. Denote the p-value with observed T_q to be $\Pr(T_q^* \ge T_q^{(b)})$. Then, the procedure to compute the p-value is as follows:

- Step 1: Generate iid $\mathcal{N}(0, 1)$ random variables $X_1^*, X_2^*, \ldots, X_N^*$.
- Step 2: Set $T_q^* = |X_{(q)}^*|/\{(\sum_{i=1}^{l-1} |X_{(i)}^*|^2)/(l-1)\}^{1/2}$, where $\{|X_{(1)}^*| < |X_{(2)}^*| < \cdots < |X_{(N)}^*|\}$ are the order statistics with respect to random samples $\{|X_1^*|, |X_{(2)}^*|, \ldots, |X_N^*|\}$.
- Step 3: Generate a realization of bootstrap samples $\{T_q^{(1)}, \ldots, T_q^{(B)}\}$ which satisfies the condition $T_q^{(b)} \ge T_q^*|H_q$ for $b = 1, \ldots, B$. Note that for each bootstrapping, the set of wavelet coefficients are re-sampled from the original set of wavelet coefficients with the same length of data points. Using the extracted distribution of $\tilde{\mathbf{T}}_q = \{\tilde{T}_q^{(1)}, \ldots, \tilde{T}_q^{(B)}\}$, re-sample the threshold and calculate the p-value as $\Pr(\tilde{T}_q^{(b)} \ge \lambda_\kappa \mid H_q) \le \kappa$ for $b = 1, \ldots, B$, through the realized bootstrap samples.

B. DETERMINATION OF THE HYPER-PARAMETERS
To execute the proposed procedure effectively, it is necessary to determine the values of hyper-parameters; the assumed number of insignificant wavelet coefficients (l), false discovery error rate (κ), and the number of simulation replications (B). In this section, we present some guidelines for selecting these hyper-parameters.

Actually, the l value should be less than the anticipated number of insignificant coefficients in order to safeguard against eliminating significant coefficients from the WSDT

procedure. In practice, if the number of extracted significant coefficients is equal or larger than the number of tested coefficients in process, we decrease the value of the lower bound l and implement the procedure again. In other words, to be self-consistent, the WSDT procedure, along with certain l, should be able to recognize the insignificant coefficients for shrinkage purpose.

Alternatively, to determine l, denote a normalized energy at position j as $\breve{\theta}_j^2 = \theta_j^2/\|\boldsymbol{\theta}\|^2$ and sort them so that $\{\breve{\theta}_{(1)}^2 < \breve{\theta}_{(2)}^2 < \ldots < \breve{\theta}_{(N)}^2\}$. Using a cumulative ordered normalized energy for the smallest value of lower bound l, $E_\theta(l) = \sum_{j=1}^{l} \breve{\theta}_{(j)}^2/\|\boldsymbol{\theta}\|^2$, the criterion for selecting the value of l is defined as

$$l = N - \sum_{j=1}^{N} \mathbb{I}(E_\theta(l) \ge (1 - \delta)), \quad (22)$$

where $(1 - \delta)$ is the energy cut-off point for a cumulative energy level of interest, δ. Because wavelet coefficients have the sparseness property, much greater cumulative normalized energies should be compacted into fewer wavelet coefficients. According to (22), we need to search for the δ values to determine the best l value under a variety of scenarios. A pictorial explanation for determining the lower bound is given in Figure 1.
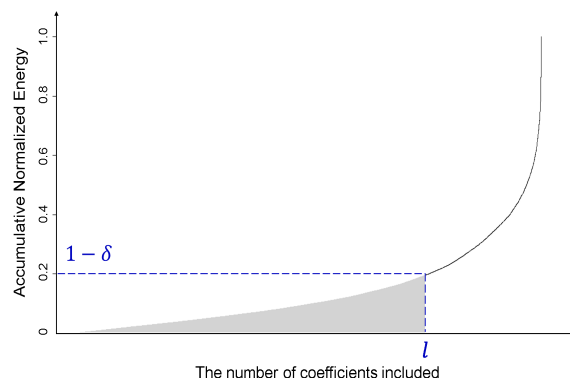

FIGURE 1. Pictorial explanation for determining a cut-off value of *l*.

The choice of κ affects the threshold level, therefore, various values of κ are implemented to investigate their impacts on the extracted wavelet features. We fix κ to either 0.01, 0.05, or 0.1 to secure 99%, 95%, or 90% confidence level in the analysis of both simulations and a practical example. In the threshold determination, user can decide the level of noise removal by adjusting hyper-parameters since the parameters, l and λ_κ, are derived by a selection of κ and δ. The number of bootstrapping replications can also affect overall thresholding performance. One way to guide against biased approximation results is to try with large value of B. In this work, we set $B = 10,000$ for the simulations and a real data analysis to achieve desirable approximation accuracy with consistent reconstruction results.

IV. APPLICATIONS AND COMPARISONS

In this section, we compare the performance of the proposed thresholding procedure with those of other existing thresholding methods by analyzing various simulated data and a real-world example. The simulated patterns in use are four well-known signals [9]: Bumps, HeaviSine, Blocks, and Doppler. For comparison, we added Gaussian white noises to the four signals to examine noise effects on the thresholding performance. In case of a real-world dataset, the sound signals obtained from an air-conditioner were used to evaluate the performance of various thresholding methods. Hereinafter, all of both soft and hard thresholdings are denoted by the subscripts '*s*' and '*h*', respectively. For the wavelet transform method, the complex Daubechies wavelet transform was used to convert original signals into wavelet-domain signals, where the family of Lina-Mayrand 3.1 wavelets (also called as Lawton's complex-valued wavelets) [16] has been applied. As a performance measure of the thresholding rules, the goodness-of-fit statistics were evaluated for each signal data in terms of average mean-square error: $AMSE = \sum_{i=1}^{N}(f_i - \hat{f}_i)^2/N$, where f_i and $\hat{f}_i$ are a true signal and an estimated (denoising) signal, respectively. For the sound data of an air-conditioner, the number of significant coefficients (N_S) and the data reduction ratio (DRR) were considered as the thresholding performance measure instead of the AMSE.
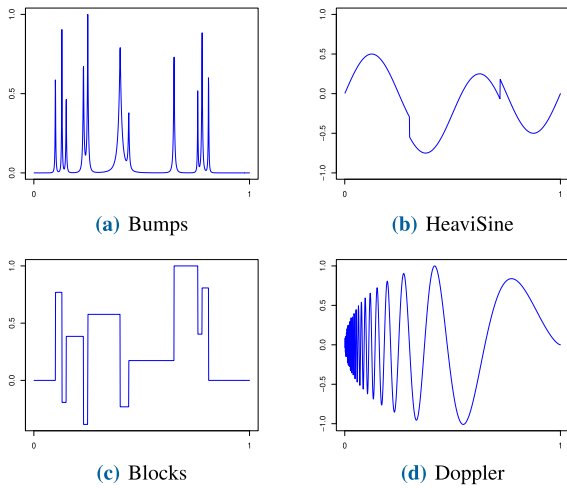

FIGURE 2. Four simulated testing signals [9].

A. SIMULATION WITH NOISY SIGNALS

Four simulated signals in Figure 2 characterize different types of features commonly observed in signal processing, seismography, manufacturing, and other engineering fields [9], [10]. Each of testing signals has $N = 1,024$ data points, containing some of fluctuations which can be observed in the real-world applications. We compared the performance of the proposed procedure in Section 3 with other existing thresholding methods using the simulated four signals.
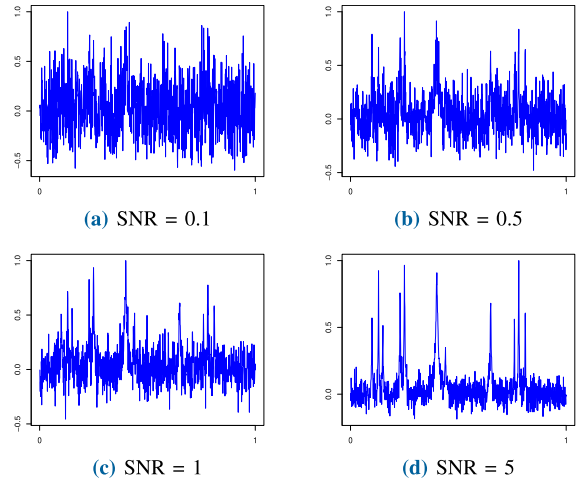

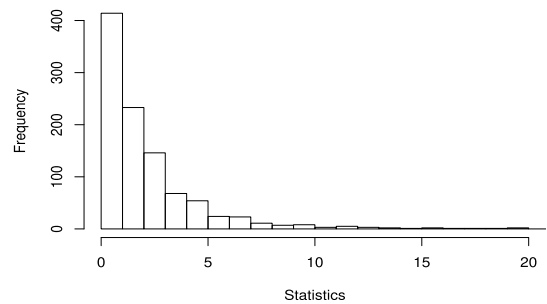**FIGURE 3.** Noisy Bumps signals at various levels of SNR.

FIGURE 4. Histogram of θ of a Bumps signal with size 2^{10} and SNR equal to 0.5.

In a series of experiments, Gaussian white noises with a zero mean were added to the four testing signals. The variance of added noises was set based on the signal to noise ratio (SNR), which decides the degrees of noises added to the original signals. For example, Figure 3 shows some noisy Bumps signals at different levels of SNR, where the smaller SNR imposes the stronger additive noises onto the original signals. For example, Figure 4 illustrates the distribution of the variance-weighted statistics $\boldsymbol{\theta}$ (at Step 1 in Section 3) for the Bumps signal of size 2^{10} with SNR equal to 0.5. Based on the variance-weighted statistics, we need to calculate the scale-invariant ratio T_q at Step 2. Figure 5 presents the T_q statistics against the assumed number of significant coefficients (q) according to various SNRs. T_q values tend to depend on the noise levels, showing that T_q value increases as the noise level increases. It is necessary to generate bootstrap samples of T_q satisfying (18) in Step 3 and calculate the p-values of the WSDT procedure in Step 4. However, because the calculation procedure of the p-values using the bootstrap samples is computationally intensive, we introduced approximation method for computing the p-values. Figure 6 shows bootstrap statistics of T_q's (solid line) against q for the Bumps signal of size 2^{10} with SNR = 0.5, along with the simulated T_q's (dashed line), which are generated from iid $\mathcal{N}(0, 1)$ by following the approximation procedure. We calculated the

TABLE 1. Average mean-square errors (AMSEs) for four testing signals using various thresholding rules.

| Function	SNR	VisuShrink$_h$	VisuShrink$_s$	CMWS$_h$	CMWS$_s$	WSDT$_h$	WSDT$_s$
Bumps	0.1	0.4822	0.5050	0.4359	0.4043	0.4240	0.3863
	0.25	0.3893	0.4589	0.2907	0.2962	0.2903	0.2442
	0.5	0.2921	0.4027	0.1773	0.1947	0.1711	0.1434
	1	0.1782	0.3243	0.0914	0.1058	0.0896	0.0788
HeaviSine	0.1	1.1802	4.2884	1.2528	0.7063	0.5749	0.6643
	0.25	0.7014	2.2681	0.6935	0.4531	0.4161	0.4336
	0.5	0.3757	1.3607	0.4640	0.3254	0.3307	0.3749
	1	0.3217	0.8576	0.2986	0.2258	0.2734	0.3496
Blocks	0.1	2.8148	4.1066	2.3850	2.2095	2.3553	2.2304
	0.25	2.0561	3.1216	1.4863	1.4931	1.7011	1.3961
	0.5	1.3792	2.4475	0.8641	0.9307	0.9796	0.8188
	1	0.7921	1.7813	0.4804	0.5294	0.5402	0.5051
Doppler	0.1	0.0620	0.0796	0.0336	0.0352	0.0399	0.0343
	0.25	0.0243	0.0571	0.0158	0.0155	0.0178	0.0174
	0.5	0.0138	0.0381	0.0112	0.0105	0.0111	0.0108
	1	0.0090	0.0247	0.0072	0.0070	0.0082	0.0068
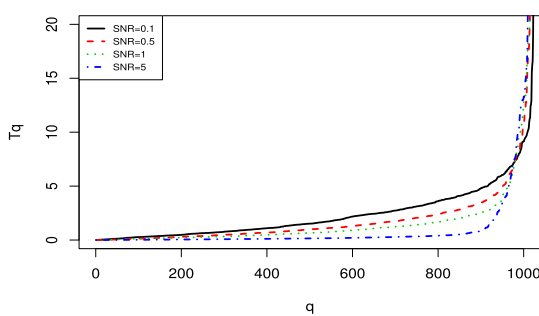
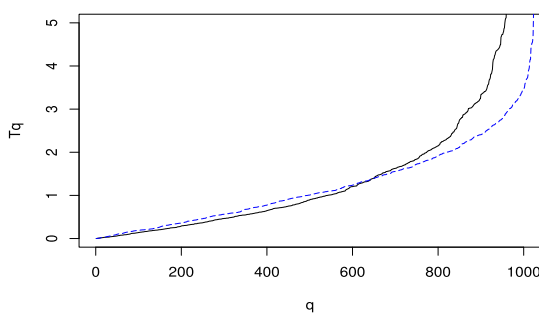**FIGURE 5.** Plots of T_q for statistics of Bumps signals depending on SNR.

FIGURE 6. Bootstrap statistics of T_q's against q for the Bumps signal of size 2^{10} with SNR = 0.5, along with Monte-Carlo simulations following iid $\mathcal{N}(0, 1)$.

(a) VisuShrink$_h$ **(b)** VisuShrink$_s$

(c) CMWS$_h$ **(d)** CMWS$_s$

(e) WSDT$_h$ **(f)** WSDT$_s$

FIGURE 7. Reconstruction of the Bumps noisy signal with SNR = 1.

maximum value of T_q satisfying $\tilde{T}_q \geq T_q|H_q$ (the crossing point of two curves in Figure 6) to consist of bootstap samples. Then, we repeated this bootstrap procedure for *B* replicates. Constructing the bootstrap samples in (18) with $B = 10,000$ replicates, that is, $\tilde{\mathbf{T}}_q \equiv \{\tilde{T}_q^{(1)}, \ldots, \tilde{T}_q^{(10000)}\}$, the *p*-value satisfying (19) will be finally calculated.

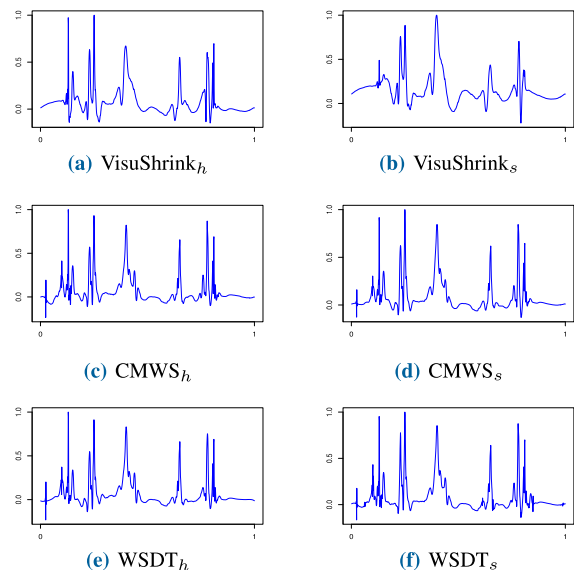Figure 7 shows reconstructed Bumps signals with SNR = 1 using various thresholding rules. The visualization results show that the VisuShrink thresholding rule provides smoother results than other thresholding rules, removing considerable amounts of fluctuations from the Bump signals. On the other hand, the CMWS and the WSDT thresholding rules perform better in that the abrupt changes in the original signals are successfully preserved. Note that the WSDT thresholding rule shows more accurate reconstruction results in the straight section. For a qualitative evaluation of shrinkage performance, Table 1 gives a summary of the performances for the four noisy signals, along with four SNR cases. Overall, AMSE tends to increase as the SNR decreases, indicating that the denoising capabilities depend on the noise degrees in true signals. AMSEs from both soft and hard WSDT thresholding rules show better reconstruction capabilities, indicating that the WSDT procedure can be successfully used for data-denoising purpose.

TABLE 2. Optimal hyper-parameter values for four testing signals.

Function	SNR	δ_h	κ_h	δ_s	κ_s	SNR	δ_h	κ_h	δ_s	κ_s
Bumps	0.1	0.01	0.01	0.26	-	0.25	0.45	-	0.64	-
	0.5	0.77	-	0.85	-	1	0.92	-	0.95	-
HeaviSine	0.1	0.20	-	0.49	-	0.25	0.44	-	0.86	-
	0.5	0.55	-	0.98	-	1	0.99	-	0.99	-
Blocks	0.1	0.19	-	0.43	-	0.25	0.69	-	0.82	-
	0.5	0.92	-	0.94	-	1	0.98	-	0.98	-
Doppler	0.1	0.29	-	0.47	-	0.25	0.69	-	0.79	-
	0.5	0.90	-	0.93	-	1	0.97	-	0.98	-

In the WSDT procedure, the hyper-parameters need to be set to efficiently threshold the wavelet coefficients according to the degrees of the noises. Accordingly, users can flexibly control the denoising intensities by adjusting the hyper-parameter values. In general, the best combinations of the hyper-parameters should be determined and it can be done by employing a kind of best model selection procedure. In this work, a grid-search approach was adopted to determine optimal values of the two parameters δ and κ. Table 2 presents the optimal values for each hyper-parameters in the WSDT rule, where δ seems to decrease as the SNR decreases, whereas κ is insensitive to the SNR. Overall, the proposed method provides improved thresholding capabilities by effectively removing interruptive noises from original signals while capturing the characteristics of original signals.

B. AN EXAMPLE: AIR-CONDITIONER REFRIGERANT NOISES

In this section, we applied the WSDT procedure to air-conditioner refrigerant noises, along with existing wavelet thresholding methods for comparison. Refrigerant noise from air-conditioners is a chronic complaining issue to users. It is crucial to discriminate the noise types in advance to eliminate or reduce the harsh noises. To classify the refrigerant noises, a microphone was used to record the refrigerant flow noises from an air-conditioner of wall-mounted type. The refrigerant noises were measured in an anechoic room with a background noise level of 22.8 decibel (dB) at the height of 0.8 m wall. It was operated over 600 seconds under cooling conditions to generate a sufficient amount of noises at 1.0 m distance from the microphone. A jury test by experts was also conducted to acoustically discriminate the refrigerant noises, where the details on the scheme were given in Cha *et al.* [7].

As a result, three types of refrigerant noises with unique audible characteristics were detected. The first noise (named ''Sound 1'') was characterized as waveforms, slugs and bubbles. The second noise (named ''Sound 2'') sounds like water drops, and the third noise (named ''Sound 3'') sounds like single-phase gas flows. The samples of three sounds are presented in Figure 8. The signals do not show any significant differences in original scale, thus we performed the WSDT to distinguish each sound by eliminating interruptive noises. In performing the WSDT procedure which considers the
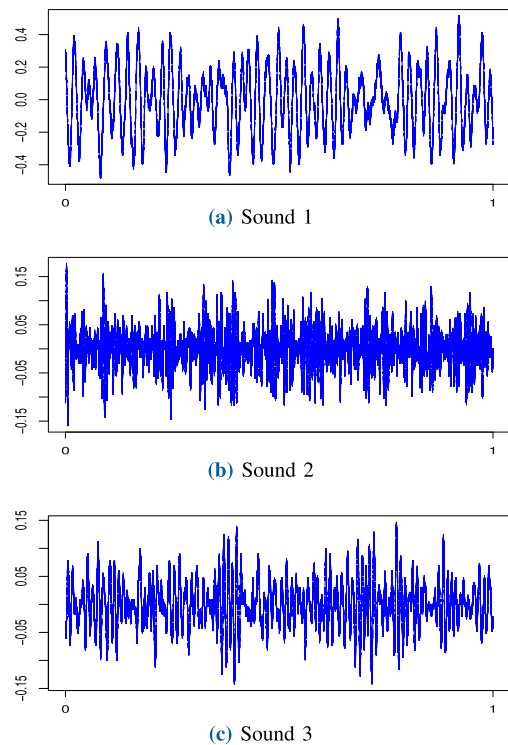

(a) Sound 1

(b) Sound 2

(c) Sound 3

FIGURE 8. Three refrigerant noise sound data from a wall-mounted air-conditioner.

difference of data size, the resolution levels were decomposed into 17, 18, and 19 levels for measurements of sound 1, sound 2, and sound 3 refrigerant noises, respectively. Instead of AMSE criterion, we used the number of significant coefficients (N_s) and data reduction ratio (DRR) as denoising performance criteria after performing thresholding rules because information about true signals was not given. The values of hyper-parameters, δ and κ, were fixed to 0.99 and 0.01.

For example, Figure 9 presents reconstructed curves using various (soft and hard) thresholding methods, as well as the WSDT thresholding procedures for Sound 1 signal. Figure 9 clearly shows that reconstructed signals using both $WSDT_h$ and $WSDT_s$ effectively capture sound characteristics (e.g., fluctuations or peaks) in the signal by thresholding interruptive noises. Table 3 gives a summary of thresholding performances resulting from the air-conditioner data.
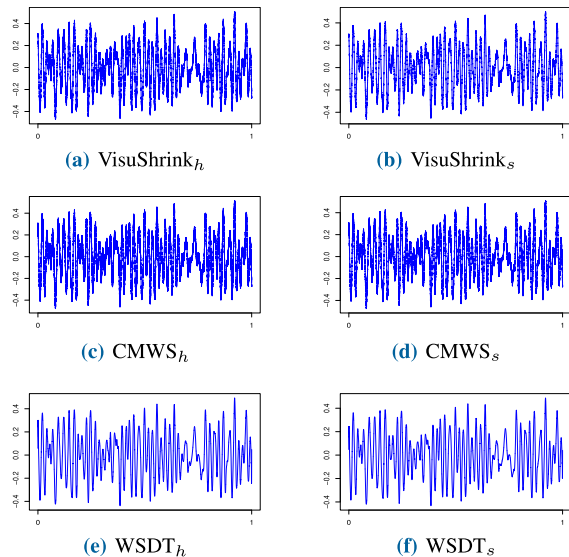
(a) VisuShrink$_h$

(b) VisuShrink$_s$

(c) CMWS$_h$

(d) CMWS$_s$

(e) WSDT$_h$

(f) WSDT$_s$

FIGURE 9. Reconstruction of the Sound 1 signal.

TABLE 3. The number of significant coefficients (N_s) and data reduction ratio (DRR) for the air-conditioner.

Type	Measures	VisuShrink	CMWS	WSDT
Sound 1	N_s	9378	20877	271
	DRR	92.85	84.07	99.79
Sound 2	N_s	32847	80220	924
	DRR	93.73	84.70	99.82
Sound 3	N_s	15046	39888	490
	DRR	94.26	84.78	99.81

Note that the WSDT procedure drastically reduces data dimension, providing smoother signals than other thresholding methods. Considering both visualization results and data reduction capabilities, the proposed theresholding procedure is effective in true signal reconstruction by removing the noises from original signals.

V. CONCLUDING REMARKS

In this paper, we propose a new wavelet thresholding procedure for data-denoising purpose. The proposed procedure is constructed based on the step-down testing procedure which efficiently identifies significant effects in unreplicated fractional factorial experiments by controlling a false discovery error rate. By testing the significance of variance-weighted statistics with boostrapping, the wavelet step-down procedure remove insignificant wavelet coefficients as noises. Performance measures of the proposed procedure are compared with other existing thresholding methods using four popular signals and a real practical example. The results show that the wavelet step-down procedure has a potential in signal-denoising and signal-reconstruction problems by efficiently thresholding interruptive noises from original signals. In particular, the WSDT procedure shows better performance in denoising and reconstruction of the signal data with

irregular variation, where the data reduction rate is flexibly adjustable according to the selection of hyper-parameters. Moreover, the proposed method effectively removes inherent noises from the data with large sample size, showing its robustness of data length and magnitude. For this purpose, it's application area can be extended to change-point problems or anomaly detections of time series data. Admittedly, the WSDT procedure depends on the hyper-parameters which control the degrees of shrinkage by adjusting false discovery error rate. In future work, the best values of the hyper-parameters can be searched using numerical methods or optimization procedures such as a Bayesian optimization.

REFERENCES
[1] F. Abramovich and Y. Benjamini, "Thresholding of wavelet coefficients as multiple hypotheses testing procedure," *Lect. Notes Statist.*, vol. 103, pp. 5–14, Apr. 1995.
[2] F. Abramovich and Y. Benjamini, "Adaptive thresholding of wavelet coefficients," *Comput. Statist. Data Anal.*, vol. 22, no. 4, pp. 351–361, Aug. 1996.
[3] F. Abramovich, T. Sapatinas, and B. W. Silverman, "Wavelet thresholding via a Bayesian approach," *J. Roy. Stat. Soc., Ser. B Stat. Methodol.*, vol. 60, no. 4, pp. 725–749, Nov. 1998.
[4] S. Barber and G. P. Nason, "Real nonparametric regression using complex wavelets," *J. Roy. Stat. Soc., Ser. B Stat. Methodol.*, vol. 66, no. 4, pp. 927–939, Nov. 2004.
[5] Y. Benjamini and Y. Hochberg, "Controlling the false discovery rate: A practical and powerful approach to multiple testing," *J. Roy. Stat. Soc., Ser. B Methodol.*, vol. 57, no. 1, pp. 289–300, Jan. 1995.
[6] E. J. Candes, C. A. Sing-Long, and J. D. Trzasko, "Unbiased risk estimates for singular value thresholding and spectral estimators," *IEEE Trans. Signal Process.*, vol. 61, no. 19, pp. 4643–4657, Oct. 2013.
[7] K. J. Cha, K.-H. Yoo, C. U. Lee, B. M. Mun, and S. J. Bae, "Classification of acoustic noise signals using wavelet spectrum based support vector machine," *J. Mech. Sci. Technol.*, vol. 32, no. 6, pp. 2453–2462, Jun. 2018.
[8] D. Clonda, J.-M. Lina, and B. Goulard, "Complex daubechies wavelets: Properties and statistical image modelling," *Signal Process.*, vol. 84, no. 1, pp. 1–23, Jan. 2004.
[9] D. L. Donoho and I. M. Johnstone, "Ideal spatial adaptation by wavelet shrinkage," *Biometrika*, vol. 81, no. 3, pp. 425–455, Sep. 1994.
[10] D. L. Donoho, "De-noising by soft-thresholding," *IEEE Trans. Inf. Theory*, vol. 41, no. 3, pp. 613–627, May 1995.
[11] D. L. Donoho and I. M. Johnstone, "Adapting to unknown smoothness via wavelet shrinkage," *J. Amer. Stat. Assoc.*, vol. 90, no. 432, pp. 1200–1224, Dec. 1995.
[12] F. Auger and P. Flandrin, "Improving the readability of time-frequency and time-scale representations by the reassignment method," *IEEE Trans. Signal Process.*, vol. 43, no. 5, pp. 1068–1089, May 1995.
[13] R. A. Gopinath, "The phaselet transform–An integral redundancy nearly shift-invariant wavelet transform," *IEEE Trans. Signal Process.*, vol. 51, no. 7, pp. 1792–1805, Jul. 2003.
[14] M. K. Jeong, J.-C. Lu, X. Huo, B. Vidakovic, and D. Chen, "Wavelet-based data reduction techniques for process fault detection," *Technometrics*, vol. 48, no. 1, pp. 26–40, Feb. 2006.
[15] J. Josse and S. Sardy, "Adaptive shrinkage of singular values," *Statist. Comput.*, vol. 26, no. 3, pp. 715–724, May 2016.
[16] J.-M. Lina and M. Mayrand, "Complex daubechies wavelets," *Appl. Comput. Harmon. Anal.*, vol. 2, no. 3, pp. 219–229, Jul. 1995.
[17] S. G. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 11, no. 7, pp. 674–693, Jul. 1989.
[18] D. C. Montgomery, *Design and Analysis of Experiments*, 6th ed. Hoboken, NJ, USA: Wiley, 2005.

[19] K. Morettin, "Wavelets in statistics," *Resenhas*, vol. 3, no. 2, pp. 211–272, 1997.

[20] S. Sardy, "Minimax threshold for denoising complex signals with waveshrink," *IEEE Trans. Signal Process.*, vol. 48, no. 4, pp. 1023–1028, Apr. 2000.

[21] G. Strang, "Wavelets and dilation equations: A brief introduction," *SIAM Rev.*, vol. 31, no. 4, pp. 614–627, Dec. 1989.

[22] J. H. Venter and S. J. Steel, "Identifying active contrasts by stepwise testing," *Technometrics*, vol. 40, no. 4, pp. 304–313, Nov. 1998.

[23] B. Vidakovic, "Nonlinear wavelet shrinkage with bayes rules and bayes factors," *J. Amer. Stat. Assoc.*, vol. 93, no. 441, pp. 173–179, Mar. 1998.

[24] B. Vidakovic, *Statistical Modeling by Wavelets*. New York, NY, USA: Wiley, 1999.

OLUFEMI A. OMITAOMU (Senior Member, IEEE) received the B.Sc. degree in mechanical engineering from Lagos State University, Nigeria, the M.Sc. degree in mechanical engineering from the University of Lagos, Nigeria, and the Ph.D. degree in information engineering from the University of Tennessee, USA. He is currently a Senior Research and Development Staff with the Oak Ridge National Laboratory, Computational Sciences and Engineering Division, Modeling and Simulation Group, Oak Ridge, TN, USA. He is also a Joint ORNL-UT Associate Professor with the Department of Industrial and Systems Engineering, University of Tennessee. His research interests include artificial intelligence in energy systems, anomaly detection in complex systems, and cognitive coupling of human-machine interactions. He is a Senior Member of the Institute of Industrial and Systems Engineers (IISE).

MUNWON LIM is currently pursuing the Ph.D. degree in industrial engineering with Hanyang University, Seoul, South Korea. Her current research interests include anomaly detection of large complex systems based on signal and image processing, prognostics, and health management of big-sized sensor data.

SUK JOO BAE (Member, IEEE) received the Ph.D. degree from the School of Industrial and Systems Engineering, Georgia Institute of Technology, in 2003. From 1996 to 1999, he has worked as a Reliability Engineer with Samsung SDI, South Korea. He is currently a Professor with the Department of Industrial Engineering, Hanyang University, Seoul, South Korea. His research interests include reliability evaluation of light displays and nano devices via accelerated life and degradation testing, statistical robust parameter design, and process control for large-volume on-line processing data. He is a member of the INFORMS, ASA, and IMS. He is also an Associate Editor of the IEEE Transactions on Reliability.

• • •