Noise Invalidation Denoising

SoosanBeheshti, Senior Member, IEEE, MasoudHashemi, Student Member, IEEE, and Xiao-PingZhang, Senior Member, IEEE, NimaNikvand, Student Member, IEEE.

Abstract—A denoising technique based on noise invalidation is proposed. The adaptive approach derives a noise signature from the noise order statistics and utilizes the signature to denoise the data. The novelty of this approach is in presenting a general-purpose denoising in the sense that it does not need to employ any particular assumption on the structure of the noise-free signal, such as data smoothness or sparsity of the coefficients. An advantage of the method is in denoising the corrupted data in any complete basis transformation (orthogonal or non-orthogonal). Experimental results show that the proposed method, called Noise Invalidation Denoising (NIDe), outperforms existing denoising approaches in terms of Mean Square Error (MSE).

Index Terms—Thresholding, order statistics, confidence region.

I. INTRODUCTION

Data denoising approaches are well known methods with presence in diverse applications and have been studied and developed in research areas ranging from communications to biomedical signal analysis. In denoising techniques, multi-resolution representations of the data is generally used. For example, wavelet shrinkage is based on rejecting those wavelet coefficients that are smaller than a certain value and keeping the remaining coefficients. Thus, the problem of removing noise from a set of observed data is transformed into finding a proper threshold for the data coefficients. The pioneer shrinkage methods, such as VisuShrink and SureShrink, propose thresholds that are functions of the noise variance and the data length [1], [2]. Over the past fifteen years, several thresholding approaches such as [3]–[7] have been developed. These methods provide optimum thresholds by focusing on certain properties of the noise-free signal, and they are proposed for particular applications, mostly for the purpose of image denoising. Unlike these approaches, the method presented in this paper focuses only on the properties of the additive noise. By relying on the noise statistics, the method defines a probabilistic region of confidence for the noise coefficients. Consequently, it validates those observed coefficients that are out of the noise confidence region and contain noiseless dominant parts. In the presence of additive colored noise, the required data preprocessing and/or whitening filters may damage the desired noiseless data. Dealing with the colored noise without whitening is especially critical if the noise correlation is due to the non-orthogonality of the basis. The proposed method denoises such data by using only the colored noise statistics in the invalidation step. The principle behind the proposed approach is simple yet powerful as it takes advantage of the properties for the additive noise and demonstrates efficiency in applications for general-purpose denoising.

The paper is organized as follows. In Section II the considered problem is formulated and revisits the philosophy of thresholding. Section III derives a noise signature from the noise statistics. Section IV presents the proposed noise invalidation approach. In Section V simulation results are provided, and conclusions are drawn in Section VI.

Notations: We use capital letters V, W, and Θ for random variables. Samples of these variables are shown with small letters v, w, and θ.

II. PROBLEM FORMULATION AND MOTIVATION

Noise-free data vector of length N, \( \tilde{y}^N = [\tilde{y}_1, \cdots, \tilde{y}_N]^T \) is corrupted by an additive white Gaussian random process \( w^N = [w_1, \cdots, w_N]^T \) with zero mean and variance of \( \sigma^2 \). The observed data \( y^N = [y_1, \cdots, y_N]^T \) is

\[
y[i] = \tilde{y}[i] + w[i]
\]

and can be expressed in terms of a desired orthonormal basis such that

\[
\theta[i] = < s_i, y^N >
\]

where \( s_i \) is an element of orthonormal basis \( S = [s_1, s_2, \cdots, s_N] \) and the desired unavailable coefficient is

\[
\hat{\theta}_i = < s_i, y^N >
\]

Therefore, the following holds

\[
\hat{y}_i = \sum_{i=1}^{N} \hat{\theta}_i s_i, \quad y^N = \sum_{i=1}^{N} \theta_i s_i
\]

and thanks to the orthonormality of the selected basis \( S \), the noise coefficients,

\[
v^N = [v_1, \cdots, v_N]^T = < S, w^N >
\]

in the available coefficients

\[
\theta_i = \hat{\theta}_i + v_i,
\]

are also independent identically distributed random variables with the same mean and variance of the noise

\[
E(V_i) = 0, \quad var(V_i) = \sigma^2 \quad 1 \leq i \leq N
\]

The observed coefficients are soft thresholded by threshold \( T_s \)

\[
\hat{\theta}_{T_s} = \begin{cases} sgn(\theta[i])(|\theta[i]| - T_s) & \text{if } |\theta[i]| \geq T_s \\ 0 & \text{Otherwise} \end{cases}
\]

Inner product of real vectors \( a \) and \( b \) is denoted by \( < a, b > = a^T b \).
The estimate of the noise-free data with this threshold is
\[ \hat{y}_{T_S}^N = \sum_{i=1}^{N} \theta_{T_S} [i] s_i \] (9)

We provide the optimum threshold by a noise invalidation process.

A. Main Idea of Thresholding

Traditionally, the main idea in thresholding is to choose a value for which all the coefficients with absolute values smaller than that value are thrown away. The rejected coefficients are the ones that we cannot fully determine their contribution to the noise-free data and therefore, we decide to dismiss them. Nevertheless, there is always a chance that some samples of the pure noise are above the threshold and some noise-free coefficients get buried among the discarded coefficients due to the additive noise. The existing thresholding methods usually focus on a class of signals, such as images, to provide a proper threshold and evaluate the quality of the methods usually focus on a class of signals, such as images, to provide a proper threshold and evaluate the quality of the thresholds based on some performance criterion such as the mean square error (MSE). Among the thresholding methods, the ones with a more general assumption on the noise-free data and therefore, we decide to dismiss them. Nevertheless, there is always a chance that some samples of the pure noise are above the threshold and some noise-free coefficients get buried among the discarded coefficients due to the additive noise. The existing thresholding methods usually focus on a class of signals, such as images, to provide a proper threshold and evaluate the quality of the thresholds based on some performance criterion such as the mean square error (MSE). Among the thresholding methods, the ones with a more general assumption on the noise-free data are VisuShrink and SureShrink, which rely on some form of piecewise smoothness of the wavelet coefficients. The rest of the thresholding methods are application oriented, for example the BayesShrink or other image denoising methods that use properties such as generalized Gaussian distribution (GGD) for the noise-free image itself. Another example is the adaptive denoising approach in [4] for neural network applications.

Here, we present a general-purpose thresholding approach that does not need to exploit any particular property of the noise-free data itself. Consequently, the thresholding approach should be invariant with respect to the order of the data. So if the data is reordered and put in an ascending order based on its absolute value, the optimum threshold should remain the same. In this case, choosing the \( m \)th coefficient of the ordered data as the threshold is equivalent to throwing away the first \( m - 1 \) coefficients.

For a general-purpose threshold, instead of concentrating on the properties of the remaining coefficients after thresholding, it is logical to focus on the dismissed coefficients. These coefficients are discarded because they are attributed to the noise or very noisy coefficients. It is rational to equivalently state that these coefficients are discarded since they behave similarly to a set of coefficients that can be generated by an associated Gaussian distribution of the additive noise. In the following section we present one of the signatures of a set that is generated by this Gaussian distribution.

Additive Noise Variance: Thresholding methods, such as SureShrink and VisuShrink, rely heavily on the value of the additive noise variance \( \sigma^2 \). In most practical applications this value is not known. The estimate of the standard deviation is usually provided by MAD approach where \( \sigma = \text{MAD} / 6.745 \),

\[ 1 \] (Probabilistic approach for finding a significant noise-free component in a coefficients is discussed in a data denoising approach in [8]. However, in this approach a Laplacian prior for a noise-free data is assumed. Preliminary work related to the proposed method is presented in [9].

III. ADDITIVE NOISE SIGNATURE

Consider the additive noise random variable \( V \) in (2) with zero mean and finite variance. Define the signature function for any value \( z \) and \( v \) as \( g(z, v) \) such that the mean and variance of this function over \( V \) are finite values:

\[ E(g(z, V)) = G_E(z) \] (10)
\[ \text{var}(g(z, V)) = G_{var}(z) \] (11)

The signature for samples of a random process of length \( N \), \( V^N = [V_1, \ldots, V_N]^T \), with IID members that have the same distribution of \( V \) is defined as

\[ g(z, v^N) = \frac{1}{N} \sum_{i=1}^{N} g(z, v_i) \] (12)

It follows that the expected value and variance of the signature are

\[ E(g(z, V^N)) = G_E(z) \] (13)
\[ \text{var}(g(z, V^N)) = \frac{1}{N} G_{var}(z) \] (14)

Detailed are shown in Appendix I. For a large data length, while the mean is a finite fixed value, the variance becomes smaller. The use of such signatures in invalidation of the additive noise is explored with the following example.

A. Signature Example: Absolute Noise Sorting (ANS)

Consider a noise signature with the following form

\[ g(z, v_i) = \begin{cases} 0 & |v_i| \leq z \\ 1 & |v_i| > z \end{cases} \] (15)

In this case, for the signature of the IID random process \( V^N \) in (12) we have [10]–[12]:

\[ E(g(z, V^N)) = F(z) \]
\[ \text{var}(g(z, V^N)) = \frac{1}{N} F(z)(1 - F(z)) \]

where \( F(\cdot) \) is the cdf of absolute value of the additive noise

\[ F(z) = 2\phi \left( \frac{z}{\sigma} \right) - 1 \] (16)

Details are provided in Appendix II. For each \( z \), the value of \( g(z, v^N) = m/N \) where \( m \) is the number of samples of \( v^N \) with absolute values smaller than \( z \). Equivalently, when sorting \( v^N \), the \( m \)th value is the largest \( v \) that is smaller than \( z \). Therefore, when sorting the \( v_8 \), the index is \( N g(z, v^N) = m \). Figure III illustrates the effect of sorting and the role of the small variance in providing a noise signature. The figure shows the behavior of 100 samples of Gaussian noise with unit variance and length 2048. As the top figure shows, with a very high probability, the values of this data are bounded between \( \pm 3\sigma \). However, if we sort the same data based on its absolute value in the middle figure,
Please provide the extracted text from the image so I can read it naturally and convert it into a plain text representation.
which according to (23) and (24) has the following mean and variance

$$E g(z, \Theta^N) = \frac{1}{N} \sum_{i=1}^{N} H(z, \bar{\theta}_i)$$  \hspace{1cm} (27)

$$\text{var}(g(z, \Theta^N)) = \frac{1}{N^2} \sum_{i=1}^{N} H(z, \bar{\theta}_i)(1 - H(z, \bar{\theta}_i))$$  \hspace{1cm} (28)

Since the value of $H(z, \bar{\theta}_i)$ in (25) is bounded between zero and one, the variance of this value is much less than its mean for large values of $N$. Therefore, a dense area will cover the sorted data with a high probability. This area becomes distinguished from the area covered by the sorted noise-only signal as the value of $z$ grows and as the nonzero coefficients become effective. This performance is illustrated in Figure 5 which shows the area covered by the sorted noisy data for Blocks signal for when SNR=5. The figure also shows the behavior of the sorted noise-only data. As it can be seen with probability 0.999997 there is no overlap between the sorted noise and sorted noisy data after a certain value of $z$.

A. Noise Invalidation in Application

Using the noise sorting signature, it is possible to invalidate the noisy coefficients with a high confidence. Figure 6 shows the application of the method. The confidence region for the noise-only data is available upon knowing or estimating the noise variance. As the sorted absolute noisy data leaves the noise confidence region, it assures that the coefficients are becoming more effective than the noisy part of the data. The largest $z$ value for which the departure occurs is the optimum threshold ($T^*$) for the noise validation problem

$$T^* = \max_x \forall z \leq x : L_N(x) \not\ni g(x, \theta^N) \not\ni U_N(x)$$ \hspace{1cm} (29)

B. Colored Noise and Thresholding

**Corollary:** If the additive noise is colored, the expected value of $g(z, V^N)$ and $g(z, \Theta^N)$ will remain the same as these expected values for the white noise in (16) and (24). For the variance of the sorted noise we have

$$\text{var}(g(z, V^N)) = \frac{1}{N} F(z)(1 - F(z)) +$$

$$\frac{1}{N^2} \sum_{i=1,j=1, i \neq j}^{N} \text{cov}(g(z, V_i)g(z, V_j))$$ \hspace{1cm} (30)

$$\leq \frac{1}{N} F(z)(1 - F(z)) +$$

$$\frac{1}{N^2} \sum_{i=1,j=1, i \neq j}^{N} [F\left(\frac{z}{\sqrt{1 + \rho_{ij}}}\right)F\left(\frac{z}{\sqrt{1 - \rho_{ij}}}\right) - F^2(z)]$$ \hspace{1cm} (31)

where

$$\rho_{ij} = \frac{R_{yy}(i-j)}{R_{yy}(0)}$$ \hspace{1cm} (32)
with $R_{vv}(0) = \sigma^2$.

Proof: In Appendix IV

The variance for the sorted noisy data is also provided in Appendix IV. As the variance indicates, the wider is the autocorrelation of the noise process with itself, the wider is the signature region of the noise and noisy data and therefore, as it is expected, it may become more difficult to distinguish the data from the noise.

V. SIMULATION RESULTS

We perform our denoising methods on noisy versions of six standard signals, Blocks, Mishmash, Bumps, and Quadchirp which are the test signals introduced in [1]. Five level decomposition with Haar wavelet is chosen for this experiment. The confidence probability of the methods for noise invalidation region is 0.999997. Figure 7 shows the six signals and their coefficients. As this figure confirms, the test signals represent a wide range of possible coefficient structures. For example Figure 8 shows the coefficient distribution of some of these signals. While signals such as Blocks have very few nonzero coefficients and many coefficients close to zero, signals such as mishmash have more uniformly distributed coefficients. Blocks and MishMash signals represent two extreme structures, while QuadChirp have a combined structure of both of these signals. We compare the proposed method with Visushrink, SureShrink which are more general-purpose thresholding approaches. On the other hand, Sure-LET and BayesShrink are image denoising methods that are performing well with the one dimensional signals. We also consider these method for comparison. We compare the performance of the methods based on their normalized reconstruction mean square error (MSE) which is

$$\frac{\| \hat{y}_r^N - \hat{y}_N \|^2}{\| \hat{y}_N \|} \quad (33)$$

where $\hat{y}_r^N$ is the resulted denoised data and $\hat{y}_N$ is the noise-free data. Table I provides the MSE of the compared methods. As the table shows, NIDe performs better than the other approaches in most of the cases.

The methods are compared for a range of SNRs both for white or colored noise. Autocorrelation of the considered colored noise is shown in Figure 9. The results of average of 100 runs are provided in Table I. It is important to note that the MSEs for all the methods have small standard deviation, much smaller than the MSE itself. As the table shows, three methods NIDe, BayesShrink and Sure-Let are comparable in presence of additive white noise. It is worth mentioning that NIDe outperforms the other two methods for the more sparse signals 90% of times for even a wider range of SNR than the range shown in the table. For the non-sparse ones such as Quad-Chirp and Mishmash, however, Sure-let performs slightly better than the other two methods with additive white noise. For the additive colored noise, NIDe is consistently outperforming the other methods.

Figures [10] show the denoised versions of Blocks with these methods for white and colored additive noise. As the figures show and Table I confirms, the denoised data with NIDe and BayesShrink in presence of a white noise are comparable while the other methods have larger MSE. In presence of colored noise, however, NIDe method outperforms the other methods.

VI. CONCLUSION

A denoising approach based on direct invalidation of the coefficients is proposed. The invalidation process uses a signature
of the additive noise in the form of a probabilistic confidence region. The signature is defined based on the statistical properties of the additive noise and is such that its standard deviation is much smaller than its mean. In this work we provided one
example of such signature which illustrates itself in simply sorting the coefficients. It was shown that such a signature represents the noise in a dense area. The density of the area depends on the noise variance and the data length. The smaller is the noise variance and/or the longer is the data length, the denser is the signature area. This will enable us to invalidate whether every coefficient is noise dominant or data dominant. The theory of the method shows its strength for any type of noise-free signal. The variance of the signature decreases as the data length grows, providing more distinguish and denser area. The method denoises in presence of not only an additive white noise, but also an additive colored noise This will also enable the user to use the proposed thresholding technique with any non-orthogonal basis. Simulation results confirmed the advantages of the proposed noise invalidation approaches in terms of the reconstruction MSE and illustrates its robust performance. While the proposed NIDe approach is a pioneer method for a general-purpose thresholding, there seems to be a great potential in further analysis, study, and expansion of such invalidation method for the purpose of denoising. For example, it is worth investigating further functions of different noise distributions that can serve as the noise signature. In addition, for the cases that the noise-free signal belongs to a particular class of signals, a potential extension would be to combine the statistical properties of the noise-free signal with the statistical structure of the additive noise in order to define noise signatures for the purpose of probabilistic invalidation.

APPENDIX I
MEAN AND VARIANCE OF THE SIGNATURE FOR IID NOISE

For the expected value of this signature from (310) we have

\[ E(g(z, V^N)) = \frac{1}{N} \sum_{i=1}^{N} E(g(z, V_i)) = G_E(z) \]  

(34)

For the variance of the signature, since \( V_i \) and \( V_j \) are independent we have

\[ E(g(z, V_i)g(z, V_j)) = E(g(z, V_i))E(g(z, V_j)) \]  

(35)

Therefore, the cross terms in the desired variance are zero and from (311) we have

\[ \text{var}(g(z, V^N)) = \frac{1}{N^2} \sum_{i=1}^{N} \text{var}(g(z, V_i)) = \frac{1}{N} G_{\text{var}}(z) \]  

(36)

APPENDIX II
MEAN AND VARIANCE OF THE SORTING SIGNATURE FOR IID NOISE

For the mean we have

\[ E(g(z, V_i)) = 1 \times Pr(|V_i| \leq z) + 0 \times Pr(|V_i| > z) \]
\[ = Pr(V_i \leq z) = F(z) \]  

(37)

Therefore,

\[ E(g(z, V^N)) = \frac{1}{N} \sum_{i=1}^{N} E(g(z, V_i)) = F(z) \]  

(38)

For the variance

\[ E(g^2(z, V_i)) = 1 \times Pr(|V_i| \leq z) + 0 \times Pr(|V_i| > z) \]
\[ = Pr(|V_i| \leq z) = F(z) \]  

(39)

From (37) and (39)

\[ \text{var}(g(z, V_i)) = E(g^2(z, V_i)) - (E(g(z, V_i)))^2 \]
\[ = F(z) - F^2(z) \]  

(40)

On the other hand

\[ E(g(z, V_i)g(z, V_j)) = 1 \times Pr(|V_i| \leq z \& |V_j| \leq z) + 0 \times Pr(|V_i| > z \lor |V_j| > z) \]
\[ = Pr(V_i \leq z) \Pr(|V_j| \leq z) = F^2(z) \]  

(41)

Therefore, we have

\[ E((g(z, V_i) - F(z))(g(z, V_j) - F(z))) = 0 \]  

(42)

which sets the cross terms in the variance of \( g(z, V) \) to zero. From (40), (42) we have

\[ \text{var}(g(z, V^N)) = \frac{1}{N^2} \sum_{i=1}^{N} \text{var}(g(z, V_i)) = \frac{1}{N} F(z)(1 - F(z)) \]  

(43)

APPENDIX III
MEAN AND VARIANCE OF THE DATA WITH SORTING SIGNATURE

For the mean we have

\[ E(g(z, \theta_i)) = Pr(|v_i + \tilde{\theta}_i| \leq z) \]
\[ Pr(z - \tilde{\theta}_i \leq v_i \leq z - \tilde{\theta}_i) = \phi\left(\frac{z - \tilde{\theta}_i}{\sigma}\right) - \phi\left(\frac{-z - \tilde{\theta}_i}{\sigma}\right) \]  

(44)

since \( \phi(-a) = 1 - \phi(a) \), we have

\[ - \phi\left(\frac{-z - \tilde{\theta}_i}{\sigma}\right) = \phi\left(\frac{z + \tilde{\theta}_i}{\sigma}\right) - 1 \]  

(45)

which results \( E(g(z, \theta_i)) \) to be the defined \( H(z, \tilde{\theta}_i) \) in (25).

For the variance, due to the structure of \( g(z, \tilde{\theta}_i) \), similar to what we had for the noise, we have

\[ E(g^2(z, \theta_i)) = E(g(z, \theta_i)) = H(z, \tilde{\theta}_i) \]  

(46)

Therefore

\[ \text{var}(g(z, \theta_i)) = E(g^2(z, \theta_i) - (E(g(z, \theta_i)))^2 \]
\[ = H(z, \tilde{\theta}_i)(1 - H(z, \tilde{\theta}_i)) \]  

(47)

Which results in the variance in (28).
APPENDIX IV
MEAN AND VARIANCE OF THE SIGNATURE FOR COLORED NOISE

In this case the autocorrelation between the zero mean Gaussian $v_i$ is denoted by $R_{vv}(m)$.

$$g(z, v^N)) = \frac{1}{N} \sum_{i=1}^{N} g(z, v_i) = \left[ \frac{1}{N} \frac{1}{N} \cdots \frac{1}{N} \right] \begin{bmatrix} g(z, v_1) \\ g(z, v_2) \\ \vdots \\ g(z, v_N) \end{bmatrix}$$

For the expected value of this function we have

$$E(g(z, V^N))) = \frac{1}{N} \sum_{i=1}^{N} E(g(z, V_i)) = F(z)$$

which is similar to that of the IID additive noise. However, for the variance, since the following holds

$$var(g(z, V^N)) = var(\frac{1}{N} \sum_{i=1}^{N} g(z, V_i)) =$$

$$\left[ \frac{1}{N} \frac{1}{N} \cdots \frac{1}{N} \right] \begin{bmatrix} var(g(z, V_1)) & cov(g(z, V_1), g(z, V_2)) & \cdots & cov(g(z, V_1), g(z, V_N)) \\ cov(g(z, V_1), g(z, V_2)) & var(g(z, V_2)) & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ cov(g(z, V_1), g(z, V_N)) & \cdots & \cdots & var(g(z, V_N)) \end{bmatrix}$$

we have

$$var(g(z, V^N))) = \frac{1}{N} F(z)(1 - F(z)) + \frac{1}{N^2} \sum_{i=1,j=1, i \neq j}^{N} cov(g(z, V_i), g(z, V_j))$$

(50)

where elements of the second term are

$$cov(g(z, V_i), g(z, V_j)) =$$

$$E(g(z, V_i)(g(z, V_i)) - E(g(z, V_i))E(g(z, V_i))$$

(51)

where the second term is simply $F^2(z)$, the first term is $Pr(|V_i| < z \& |V_j| < z)$. For the first term, the joint distribution of $V_i$ and $V_j$ is a Gaussian distribution with zero mean and variance

$$E \left[ \begin{bmatrix} V_i \\ V_j \end{bmatrix} \right] = \left[ \begin{bmatrix} 0 \\ 0 \end{bmatrix} \right], \quad var \left[ \begin{bmatrix} V_i \\ V_j \end{bmatrix} \right] = \sigma^2 \begin{bmatrix} 1 & \rho_{ij} \\ \rho_{ij} & 1 \end{bmatrix}$$

(52)

where

$$\rho_{ij} = \frac{R_{vv}(i-j)}{R_{vv}(0)}$$

(53)

with $R_{vv}(0) = \sigma^2$. The decomposition of the covariance matrix is as follows

$$\frac{1}{2} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} 1 + \rho_{ij} & 0 \\ 0 & 1 - \rho_{ij} \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

(54)

Therefore, by the following transformation

$$\begin{bmatrix} x_i \\ x_j \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} v_i \\ v_j \end{bmatrix}$$

(55)

the two $x_i$ and $x_j$ random variables are independent and with variances $\sigma^2(1 + \rho_{ij})$ and $\sigma^2(1 - \rho_{ij})$. As a result, for the first term of the covariance

$$E(g(z, V_i)(g(z, V_i)) = Pr(|V_i| < z \& |V_j| < z)$$

(56)

$$\leq Pr(|X_i| \leq \sqrt{2}z \& |X_j| \leq \sqrt{2}z)$$

(57)

$$= F(\frac{\sqrt{2}z}{1 + \rho_{ij}})F(\frac{\sqrt{2}z}{1 - \rho_{ij}})$$

(58)

Figure 11. The desired area for calculation of the probabilities in [50] and [51].

A. Noisy Data
With similar analogy, for the signal in presence of the noisy data, we have

$$E(g(z, \Theta^N))) = \frac{1}{N} \sum_{i=1}^{N} H(z, \tilde{\theta}_i)$$

(59)

where $H(z, \tilde{\theta}_i)$ was defined in [25]. For the variance of the sorted absolute value of the noisy data, similarly [50] holds. Therefore, structure of the variance is similar to [50]:

$$var(g(z, \Theta^N))) = \frac{1}{N^2} \sum_{i=1}^{N} H(z, \tilde{\theta}_i) - H^2(z, \tilde{\theta}_i) +$$

$$\frac{1}{N^2} \sum_{i=1,j=1, i \neq j}^{N} cov(g(z, \theta_i), g(z, \theta_j))$$

(60)

For the covariance in the second term

$$cov(g(z, \Theta_i), g(z, \Theta_j)) =$$

$$E(g(z, \Theta_i)(g(z, \Theta_i)) - E(g(z, \Theta_i))E(g(z, \Theta_i))$$

(61)

we use [59] to calculate $E(g(z, \Theta_i))E(g(z, \Theta_i))$ and have

$$E(g(z, \Theta_i)(g(z, \Theta_i)) = Pr(|V_i + \tilde{\theta}_i| \leq z \& |V_j + \tilde{\theta}_j| \leq z)$$

$$= Pr(\sqrt{2}(z - \tilde{\theta}_i + \tilde{\theta}_j) \leq X_i \leq \sqrt{2}(z - \tilde{\theta}_i + \tilde{\theta}_j))$$

(62)

$$\& \sqrt{2}(z - \tilde{\theta}_j - \tilde{\theta}_j) \leq X_j \leq \sqrt{2}(z - \tilde{\theta}_i - \tilde{\theta}_j))$$

(63)

$$= H(\frac{\sqrt{2}z}{\sqrt{1 + \rho_{ij}}}, \frac{\tilde{\theta}_i + \tilde{\theta}_j}{\sqrt{1 + \rho_{ij}}})H(\frac{\sqrt{2}z}{\sqrt{2}z}, \frac{\tilde{\theta}_i - \tilde{\theta}_j}{\sqrt{2}z})$$

(64)
REFERENCES

[1] D. L. Donoho and I. M. Johnstone, “Ideal spatial adaption via wavelet shrinkage,” Biometrika, vol. 81, pp. 425-455, 1994.
[2] D. L. Donoho and I. M. Johnstone, “Adapting to unknown smoothness via wavelet shrinkage,” Journal of the American Statistical Assos. vol 90, pp.1200-1224, 1995
[3] T. Blu and F. Luisier, “The SURE-LET Approach to Image Denoising,” IEEE Transactions on Image Processing, vol. 16, no. 11, pp. 2778-2786, November 2007.
[4] X-P. Zhang, “Thresholding neural network for adaptive noise reduction,” IEEE Transactions on Neural Networks, vol. 12, no. 3, pp. 567-584, May 2001.
[5] S.G. Chang, B. Yu , and M. Vetterli, “Adaptive wavelet thresholding for image denoising and compression,” IEEE Transactions on Image Processing, vol. 9, no. 9, pp. 1532-1546, Sept. 2000.
[6] S. G. Chang, B. Yu, and M. Vetterli, “Spatially adaptive wavelet thresholding with context modeling for image denoising,” IEEE Transactions on Image Processing, vol. 9, no. 9, pp. 1522-1531, Sept. 2000.
[7] F. Abramovitch, T. Sapatinas, and B.W. Silverman, “Wavelet thresholding via a Bayesian approach,” J. Roy. Statist. Soc., ser. B, vol. 60, no. 4, pp. 725-749, 1998.
[8] A. Pizurica and W. Philips, “Estimating the probability of the presence of a signal of interest in multiresolution single- and multiband image denoising”, IEEE Transactions on Image Processing, vol. 15, no. 3, pp. 654-665 , March 2006.
[9] S. Beheshti, N. Nikvand, and X.N. Fernando, “Soft thresholding by noise invalidation,” Proceedings of the 24th Biennial Symposium on Communications, pp. 235-236, 2009.
[10] A. H. David and H. N. Nagaraja, “Order Statistics.” 3rd Edition, Wiley, New Jersey, 2003.
[11] L. Wasserman, All of Nonparametric Statistics, Springer, New York, 1st Edition, 2005.
[12] Y. Blanco, S. Zazo, and J. C. Principe, “Alternative statistical Gaussianity measure using the cumulative density function”, Proceedings of the International Workshop on ICA and Blind Signal Separation, pp. 537-542, 2000.
$H(z, \bar{\theta})$ for $\bar{\theta} = 0, 5, 10, 15$.
$H(z, \theta)$

$\sigma = 1 \quad \rightarrow \quad \sigma = 6$
