

Noise Invalidation Denoising

Soosan Beheshti, *Senior Member, IEEE*, Masoud Hashemi, *Student Member, IEEE*,
Xiao-Ping Zhang, *Senior Member, IEEE*, and Nima Nikvand, *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—Confidence region, order statistics, thresholding.

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, multiresolution 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 , $\bar{y}^N = [\bar{y}_1, \dots, \bar{y}_N]^T$ is corrupted by an additive white Gaussian random process $w^N = [w_1, \dots, w_N]^T$ with zero mean and variance of σ^2 . The observed data $y^N = [y_1, \dots, y_N]^T$ is

$$y[n] = \bar{y}[n] + w[n] \quad (1)$$

and can be expressed in terms of a desired orthonormal basis such that¹:

$$\theta[i] = \langle s_i, y^N \rangle \quad (2)$$

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

$$\bar{\theta}_i = \langle s_i, \bar{y}^N \rangle. \quad (3)$$

Therefore, the following holds

$$\bar{y}^N = \sum_{i=1}^N \bar{\theta}_i s_i, \quad y^N = \sum_{i=1}^N \theta_i s_i \quad (4)$$

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

$$v^N = [v_1, \dots, v_N]^T = \langle S, w^N \rangle \quad (5)$$

¹Inner product of real vectors a and b is denoted by $\langle a, b \rangle = a^T b$.

Manuscript received June 21, 2010; accepted August 15, 2010. Date of publication September 9, 2010; date of current version November 17, 2010. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Deniz Erdogmus.

S. Beheshti, M. Hashemi, and X.-P. Zhang are with the Department of Electrical and Computer Engineering, Ryerson University, Toronto, ON, Canada M5B 2K3 (e-mail: soosan@ee.ryerson.ca; s3hashem@ee.ryerson.ca; xzhang@ee.ryerson.ca).

N. Nikvand is with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, Canada N2L 3G1 (e-mail: nnikvand@uwaterloo.ca).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TSP.2010.2074199

in the available coefficients

$$\theta_i = \bar{\theta}_i + v_i \quad (6)$$

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

$$E(V_i) = 0, \text{var}(V_i) = \sigma^2 \quad 1 \leq i \leq N. \quad (7)$$

The observed coefficients are soft thresholded by threshold T_s

$$\hat{\theta}_{T_s} = \begin{cases} \text{sgn}(\theta[i])(|\theta[i]| - T_s) & \text{if } |\theta[i]| \geq T_s \\ 0 & \text{Otherwise.} \end{cases} \quad (8)$$

The estimate of the noise-free data with this threshold is

$$\hat{y}_{T_s}^N = \sum_{i=1}^N \hat{\theta}_{T_s}[i]s_i. \quad (9)$$

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 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 distribution².

Additive Noise Variance: Thresholding methods, such as SureShrink and VisuShrink, rely heavily on the value of the additive noise variance σ^2 . In most practical applications this value is not known. The estimate of the standard deviation is usually provided by MAD approach where $\hat{\sigma} = \text{MAD}/.6745$, and where MAD is the median of absolute value of normalized fine scale wavelet coefficients [2]. The method provided in this paper also requires the estimate of noise variance, and we use the MAD method for this estimation.

III. ADDITIVE NOISE SIGNATURE

Consider the additive noise random variable V in (7) 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:

$$E(g(z, V)) = G_E(z) \quad (10)$$

$$\text{var}(g(z, V)) = G_{\text{var}}(z). \quad (11)$$

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

$$g(z, v^N) = \frac{1}{N} \sum_{i=1}^N g(z, v_i). \quad (12)$$

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

$$E(g(z, V^N)) = G_E(z) \quad (13)$$

$$\text{var}(g(z, V^N)) = \frac{1}{N} G_{\text{var}}(z). \quad (14)$$

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

$$g(z, v_i) = \begin{cases} 1 & \text{if } |v_i| \leq z \\ 0 & |v_i| > z. \end{cases} \quad (15)$$

In this case, for the signature of the i.i.d. random process V^N in (12) we have [10]–[12]

$$E(g(z, V^N)) = F(z) \quad (16)$$

$$\text{var}(g(z, V^N)) = \frac{1}{N} F(z)(1 - F(z)) \quad (17)$$

²Probabilistic 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]

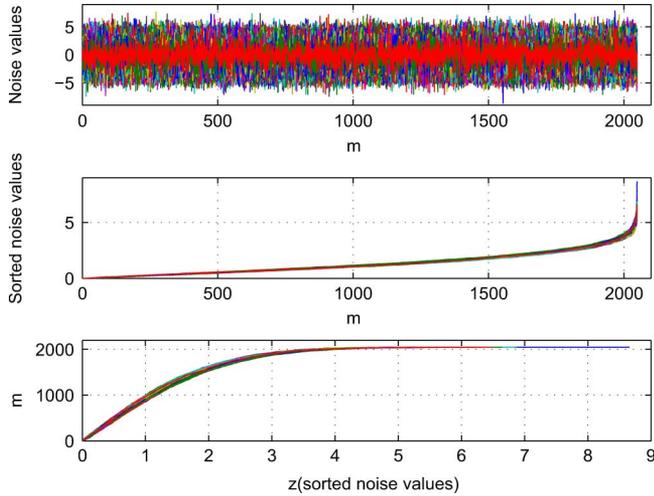


Fig. 1. Top figure: 100 runs of a zero mean Gaussian distribution with unit variance and length 2048. Middle: The same 100 runs of the above figure sorted based on their absolute values. Bottom: This is the middle figure with its vertical and horizontal axes swapped ($m = Ng(z, v^N)$).

where $F(\cdot)$ is the cumulative distribution function (cdf) of *absolute value* of the additive noise

$$F(z) = 2\phi\left(\frac{z}{\sigma}\right) - 1. \quad (18)$$

Details are provided in Appendix II. \square

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_i that is smaller than z . Therefore, when sorting the v_i s, the index is $Ng(z, v^N) = m$. Fig. 1 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, the values collapse in a much denser area. Such behavior can be explained by the ANS signature as follows. The bottom figure shows the result of swapping the horizontal and vertical axes of the middle figure. Here the horizontal axis is z and the vertical shows 100 samples of $Ng(z, v^N)$ where $N = 2048$. As it is expected, these values are around mean $NF(z)$ with variance $F(z)(1 - F(z))$. This will allow us to define proper confidence regions, with a high probability p , around the noise signature. Due to the noise signature structure, these regions are considerably smaller than the corresponding confidence regions of the Gaussian distribution of the additive noise itself. Therefore, for each z and for a high confidence probability p , we can find $L_N(z)$ and $U_N(z)$ around the mean value of $F(z)$ such that

$$Pr\{L_N(z) \leq g(z, v_N) \leq U_N(z)\} = p. \quad (19)$$

For example, Fig. 2 shows the bounds on $g(z, V^N)$ for confidence probability $p = 0.999997$ and with $\sigma = 2.5$.

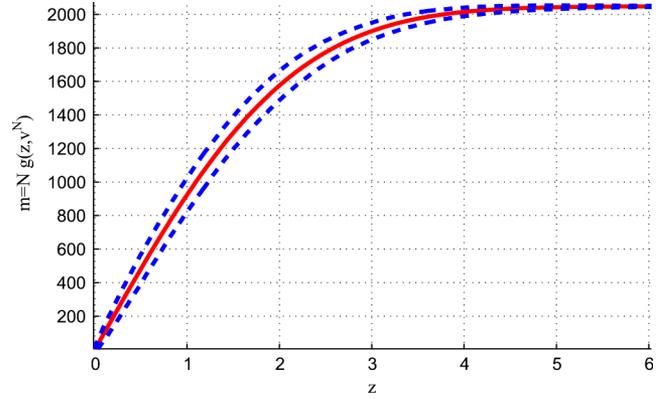


Fig. 2. Solid line: Mean of the noise $g(z, V^N)$. Dashed lines are upper and lower bounds with confidence probability 0.999997.

B. Confidence Region and Gaussian Estimate

While it is straightforward to make a table of values of the boundaries shown in Fig. 2, it is possible to use Gaussian estimates for the distributions of $g(z, v_N)$ for large enough values of N as $g(z, v_N)$ in (12) is average of N independent variables. Using the Central Limit Theorem for this distribution, we have³

$$Pr\left\{\frac{|g(z, v_N) - F(z)|}{\lambda\sqrt{\frac{1}{N}F(z)(1 - F(z))}} \leq 1\right\} \approx \text{erf}\left(\frac{\lambda}{\sqrt{2}}\right). \quad (21)$$

This estimates the boundaries in (19) to be

$$L_N(z) = F(z) - \lambda\sqrt{\frac{1}{N}F(z)(1 - F(z))} \quad (22)$$

$$U_N(z) = F(z) + \lambda\sqrt{\frac{1}{N}F(z)(1 - F(z))}. \quad (23)$$

The choice of λ should be such that the probability is close to one and at the same time the boundary is not very loose. In statistics the three-sigma rule, or empirical rule, states that for a normal distribution, almost all values lie within three standard deviations of the mean⁴. For a better quality measure, the six sigma approach increases the standard deviation to 4.5 (equivalently $p = 0.999997$). Consequently, we suggest choosing λ such that $3 \leq \lambda \leq 5$. Interestingly, our experimental observation shows that the threshold associated with $\lambda = 4.5$ provides the optimum threshold with respect to MSE in 90% of cases.

IV. NOISE INVALIDATION WITH ABSOLUTE COEFFICIENT SORTING (ACS)

The coefficients of our observed data is in form of $\theta_i = v_i + \bar{\theta}_i$ which has the same structure as the noise except its mean which

³The error function is

$$\text{erf}(x) = \frac{1}{\sqrt{\pi}} \int_0^x e^{-t^2} dt \quad (20)$$

⁴The following is known for $\lambda = 3, 4.5$ and 5

$$\text{erf}\left(\frac{3}{\sqrt{2}}\right) = 0.997300203937, \quad \text{erf}\left(\frac{4.5}{\sqrt{2}}\right) = 0.999997 \quad (24)$$

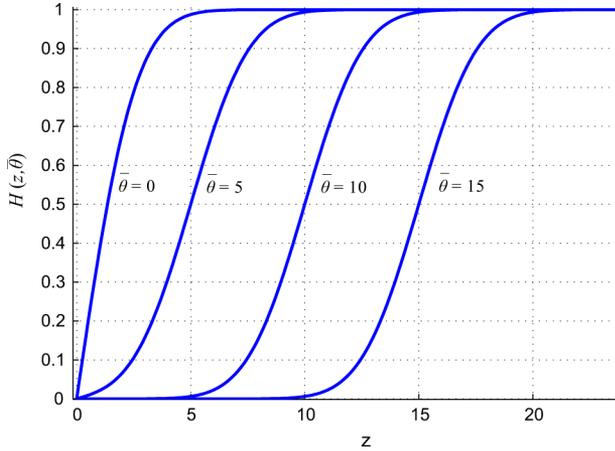


Fig. 3. Expected value of $g(z, \Theta)$ for various values of $\bar{\theta}$ when the additive noise variance is $\sigma = 4$.

is the noiseless coefficient $\bar{\theta}_i$. In this case

$$E(g(z, \Theta_i)) = Pr(|v_i + \bar{\theta}_i| \leq z) = H(z, \bar{\theta}_i) \quad (25)$$

$$var(g(z, \Theta_i)) = H(z, \bar{\theta}_i) (1 - H(z, \bar{\theta}_i)) \quad (26)$$

where

$$H(z, \bar{\theta}_i) = \phi\left(\frac{z - \bar{\theta}_i}{\sigma}\right) + \phi\left(\frac{z + \bar{\theta}_i}{\sigma}\right) - 1. \quad (27)$$

Details are provided in Appendix III. \square

Figs. 3 and 4 show typical behaviors of $H(z, \bar{\theta}_i)$ for various $\bar{\theta}_i$ s and change of noise variance. Sorting the coefficients in this case is analogous to calculation of

$$g(z, \theta^N) = \frac{1}{N} \sum_{i=1}^N g(z, \theta_i) \quad (28)$$

which according to (25) and (26) has the following mean and variance

$$E(g(z, \Theta^N)) = \frac{1}{N} \sum_{i=1}^N H(z, \bar{\theta}_i) \quad (29)$$

$$var(g(z, \Theta^N)) = \frac{1}{N^2} \sum_{i=1}^N H(z, \bar{\theta}_i) (1 - H(z, \bar{\theta}_i)). \quad (30)$$

Since the value of $H(z, \bar{\theta}_i)$ in (27) 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 Fig. 5 which shows the area covered by the sorted noisy data for Blocks signal for when $\text{SNR} = 5$. The figure also shows the behavior of the sorted noise-only data. As it can be seen with probability 0.99997 there is no overlap between the sorted noise and sorted noisy data after a certain value of z .

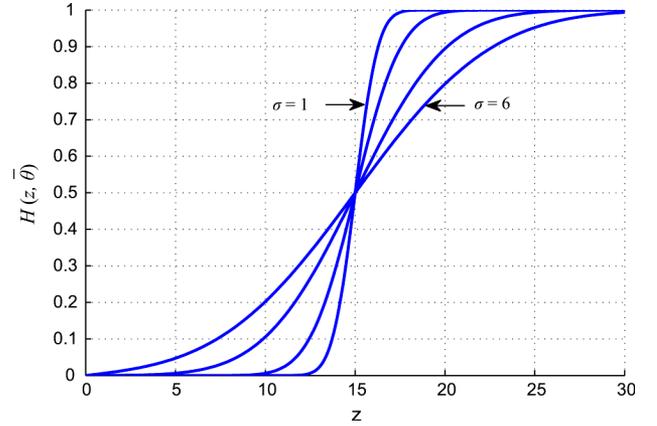


Fig. 4. Expected value of $g(z, \Theta)$ when $\bar{\theta} = 15$ and the noise standard deviations $\sigma = 1, 2, 4$ and 6 .

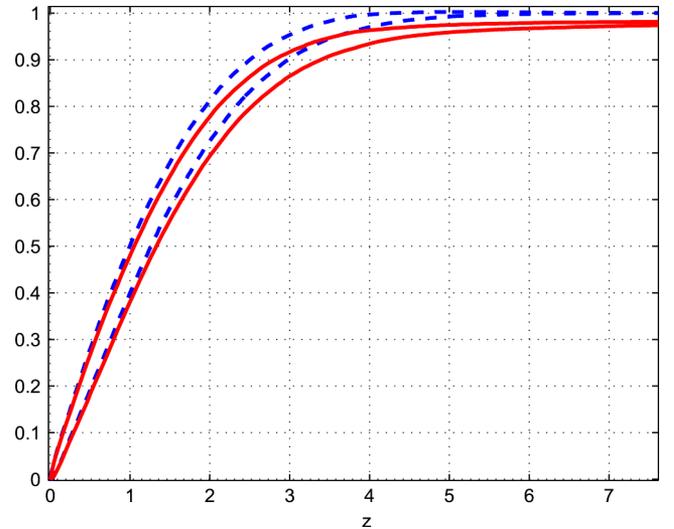


Fig. 5. The area between the solid lines is the confidence region of sorted absolute values of the noisy data coefficients of Blocks signal ($\text{SNR} = 5$) with probability 0.999997. The area between the dashed lines is the noise confidence region with probability 0.999997.

A. Noise Invalidation in Application

Using the noise sorting signature, it is possible to invalidate the noisy coefficients with a high confidence. Fig. 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

$$T^* = \max_z \forall z \leq x : L_N(x) \not\leq g(x, \theta^N) \not\leq U_N(x) \quad (31)$$

B. Colored Noise and Thresholding

Corollary 1: 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

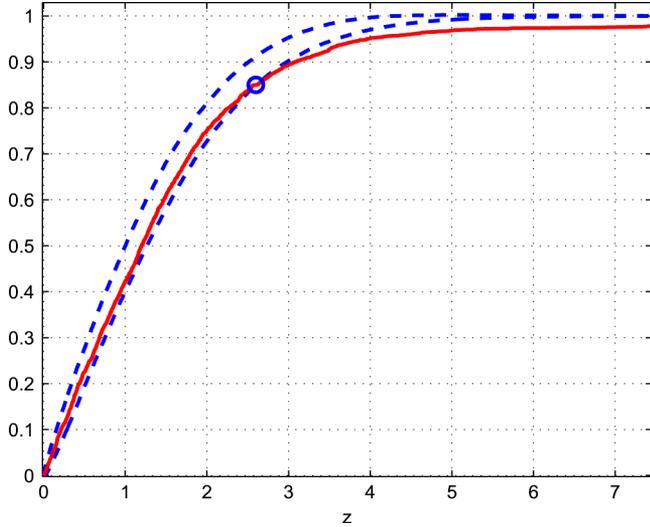


Fig. 6. Solid line is the sorted absolute values of the observed data coefficients crossing upper bound of the noise confidence region at $z = 2.6$ when the observed data is noisy Blocks signal (SNR = 5). The area between the dashed lines is the noise confidence region with probability 0.999997.

expected values for the white noise in (16) and (29). For the variance of the sorted noise we have

$$\begin{aligned} \text{var}(g(z, V^N)) &= \frac{1}{N} F(z)(1 - F(z)) \\ &+ \frac{1}{N^2} \sum_{i,j=1, i \neq j}^N \text{cov}(g(z, V_i)g(z, V_j)) \\ &\leq \frac{1}{N} F(z)(1 - F(z)) \\ &+ \frac{1}{N^2} \sum_{i,j=1, i \neq j}^N \left[F\left(\frac{z}{\sqrt{1+\rho_{ij}}}\right) F\left(\frac{z}{\sqrt{1-\rho_{ij}}}\right) - F^2(z) \right] \end{aligned} \quad (32)$$

$$(33)$$

where

$$\rho_{ij} = \frac{R_{vv}(i-j)}{R_{vv}(0)} \quad (34)$$

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. Fig. 7 shows the six signals and their coefficients. As this figure confirms, the test signals represent a wide

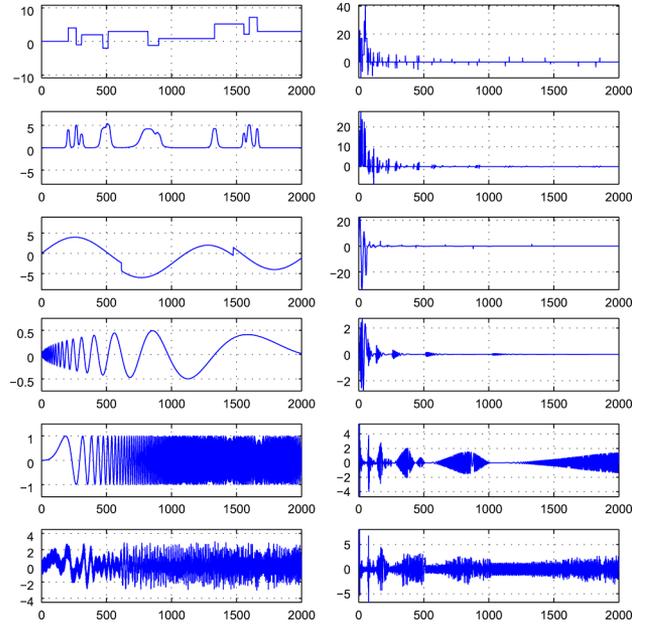


Fig. 7. From top to bottom: Blocks, Bumps, HeavySin, Doppler, QuadChirp, and MishMash. Left figures are the signals and right figures are their corresponding wavelet coefficients.

range of possible coefficient structures. For example Fig. 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 BayeshShrink 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

$$\frac{\|\hat{y}_T^N - \bar{y}^N\|^2}{\|\bar{y}^N\|^2} \quad (35)$$

where \hat{y}_T^N (9) is the resulted denoised data and \bar{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 Fig. 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 nonsparse ones such as Quad-Chirp and Mish-mash, however, Sure-let performs slightly better than the other

TABLE I
NORMALIZED RECONSTRUCTION MSE FOR THE THRESHOLDING METHODS. RIGHT FOR THE WHITE ADDITIVE NOISE AND LEFT FOR THE COLORED ADDITIVE NOISE WITH AUTOCORRELATION IN FIG. 9. AVERAGED OVER 100 RUNS

	Visu	SURE	Bayes Shrink	SURE -LET	NIDe		Visu	SURE	Bayes Shrink	SURE -LET	NIDe
Blocks						Blocks					
SNR=1	0.168	0.657	0.111	0.124	0.066	SNR=1	0.406	0.695	0.562	6.212	0.385
SNR=4	0.124	0.259	0.058	0.065	0.042	SNR=4	0.212	0.334	0.286	3.107	0.210
SNR=8	0.086	0.032	0.026	0.027	0.020	SNR=8	0.098	0.124	0.121	1.214	0.092
SNR=10	0.072	0.015	0.015	0.017	0.013	SNR=10	0.063	0.074	0.079	0.753	0.060
SNR=14	0.048	0.007	0.007	0.007	0.005	SNR=14	0.030	0.035	0.037	0.295	0.024
Bumps						Bumps					
SNR=1	0.155	0.427	0.098	0.122	0.091	SNR=1	0.455	0.648	0.559	6.443	0.427
SNR=4	0.127	0.065	0.073	0.063	0.070	SNR=4	0.279	0.305	0.293	3.220	0.245
SNR=8	0.096	0.026	0.023	0.027	0.025	SNR=8	0.105	0.115	0.125	1.273	0.098
SNR=10	0.083	0.023	0.019	0.018	0.017	SNR=10	0.072	0.075	0.081	0.773	0.072
SNR=14	0.062	0.017	0.012	0.009	0.009	SNR=14	0.035	0.038	0.035	0.299	0.030
HeavySin						HeavySin					
SNR=1	0.137	0.670	0.029	0.115	0.028	SNR=1	0.402	0.712	0.531	6.527	0.334
SNR=4	0.096	0.268	0.017	0.057	0.017	SNR=4	0.209	0.328	0.270	3.256	0.173
SNR=8	0.063	0.032	0.010	0.023	0.009	SNR=8	0.085	0.117	0.105	1.280	0.071
SNR=10	0.050	0.007	0.016	0.015	0.007	SNR=10	0.057	0.069	0.068	0.803	0.047
SNR=14	0.035	0.004	0.003	0.006	0.003	SNR=14	0.027	0.023	0.028	0.317	0.020
Doppler						Doppler					
SNR=1	0.798	0.098	0.069	0.126	0.078	SNR=1	0.446	0.399	0.569	6.431	0.401
SNR=4	0.659	0.085	0.085	0.067	0.076	SNR=4	0.330	0.236	0.293	3.201	0.235
SNR=8	0.493	0.078	0.036	0.030	0.032	SNR=8	0.231	0.138	0.127	1.207	0.110
SNR=10	0.424	0.076	0.029	0.020	0.024	SNR=10	0.187	0.113	0.082	0.76	0.075
SNR=14	0.308	0.071	0.012	0.010	0.009	SNR=14	0.133	0.089	0.037	0.038	0.030
QuadChirp						QuadChirp					
SNR=1	0.931	0.782	0.466	0.447	0.637	SNR=1	0.964	1.183	1.139	1.298	0.594
SNR=4	0.903	0.771	0.284	0.276	0.357	SNR=4	0.978	0.965	0.489	0.536	0.326
SNR=8	0.864	0.757	0.129	0.131	0.151	SNR=8	0.986	0.833	0.164	0.186	0.137
SNR=10	0.841	0.753	0.086	0.086	0.096	SNR=10	0.990	0.801	0.100	0.117	0.088
SNR=14	0.780	0.750	0.038	0.037	0.035	SNR=14	0.980	0.771	0.039	0.051	0.037
MishMash						MishMash					
SNR=1	0.926	0.523	0.517	0.462	0.693	SNR=1	0.974	1.186	1.205	1.246	0.607
SNR=4	0.906	0.430	0.318	0.286	0.374	SNR=4	0.956	0.889	0.523	0.523	0.336
SNR=8	0.865	0.432	0.146	0.136	0.156	SNR=8	0.961	0.727	0.174	0.180	0.142
SNR=10	0.836	0.623	0.095	0.089	0.099	SNR=10	0.963	0.699	0.105	0.109	0.091
SNR=14	0.752	0.641	0.040	0.039	0.039	SNR=14	0.963	0.675	0.042	0.041	0.031

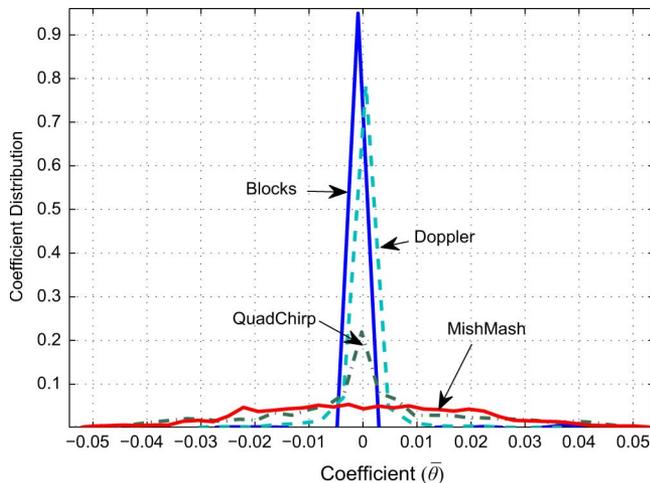


Fig. 8. Coefficient distribution for Blocks, Doppler, Quadchirp and MishMash.

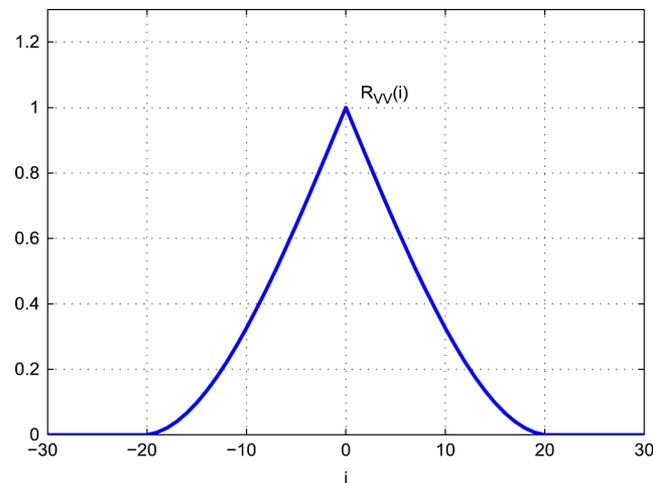


Fig. 9. Autocorrelation of the colored additive noise.

two methods with additive white noise. For the additive colored noise, NIDe is consistently outperforming the other methods.

Figs. 10 and 11 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

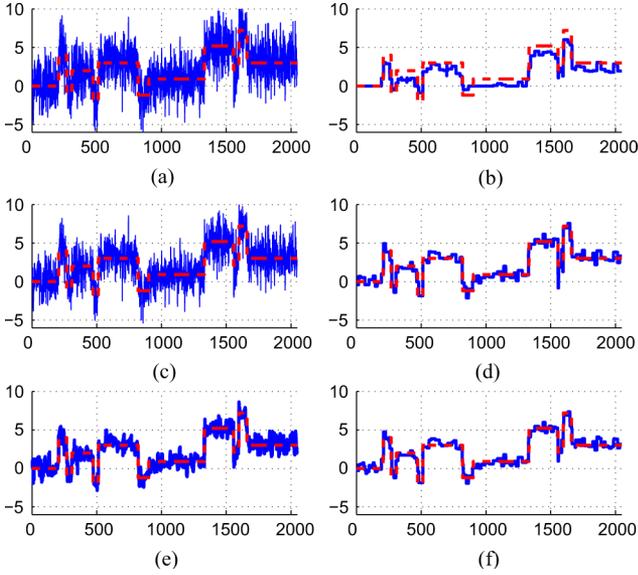


Fig. 10. Additive white noise with SNR = 4. (a) Noisy Blocks. (b) VISU. (c) SURE. (d) BayesShrink. (e) Sure-Let. (f) NIDE.

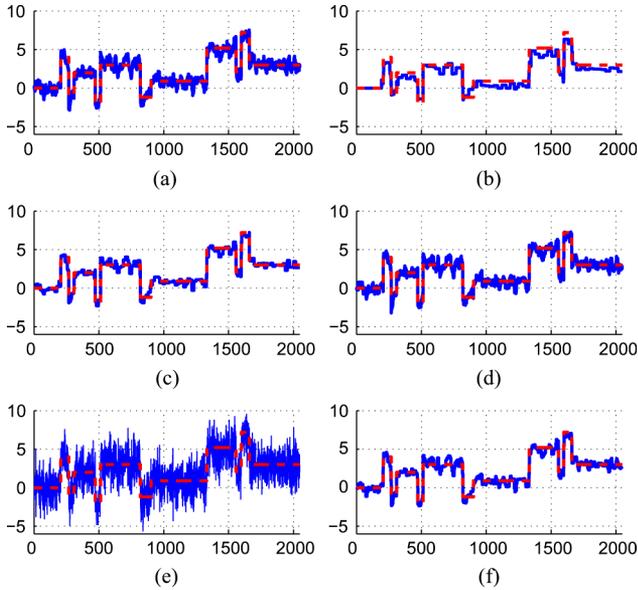


Fig. 11. Additive colored noise with SNR = 14. (a) Noisy Blocks. (b) VISU. (c) SURE. (d) BayesShrink. (e) Sure-Let. (f) NIDE.

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 (10) we have

$$E(g(z, V^N)) = \frac{1}{N} \sum_{i=1}^N E(g(z, V_i)) = G_E(z). \quad (36)$$

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

$$E(g(z, V_i)g(z, V_j)) = E(g(z, V_i))E(g(z, V_j)). \quad (37)$$

Therefore, the cross terms in the desired variance are zero and from (11) we have

$$\begin{aligned} \text{var}(g(z, V^N)) &= \frac{1}{N^2} \sum_{i=1}^N \text{var}(g(z, V_i)) \\ &= \frac{1}{N} G_{\text{var}}(z). \end{aligned} \quad (38)$$

APPENDIX II

MEAN AND VARIANCE OF THE SORTING SIGNATURE FOR I.I.D. NOISE

For the mean we have

$$\begin{aligned} E(g(z, V_i)) &= 1 \times Pr(|V_i| \leq z) + 0 \times Pr(|V_i| > z) \\ &= Pr(V_i \leq z) = F(z). \end{aligned} \quad (39)$$

Therefore

$$E(g(z, V^N)) = \frac{1}{N} \sum_{i=1}^N E(g(z, V_i)) = F(z). \quad (40)$$

For the variance

$$\begin{aligned} E(g^2(z, V_i)) &= 1 \times Pr(|V_i| \leq z) + 0 \times Pr(|V_i| > z) \\ &= Pr(|V_i| \leq z) = F(z). \end{aligned} \quad (41)$$

From (39) and (41)

$$\begin{aligned} \text{var}(g(z, V_i)) &= E(g^2(z, V_i)) - (E(g(z, V)))^2 \\ &= F(z) - F^2(z). \end{aligned} \quad (42)$$

On the other hand

$$\begin{aligned} E(g(z, V_i)g(z, V_j)) &= 1 \times Pr(|V_i| \leq z \& |V_j| \leq z) \\ &\quad + 0 \times Pr(|V_i| > z \text{ or } |V_j| > z) \\ &= Pr(V_i \leq z) Pr(|V_j| \leq z) \\ &= F^2(z). \end{aligned} \quad (43)$$

Therefore, we have

$$E((g(z, V_i) - F(z))(g(z, V_j) - F(z))) = 0 \quad (44)$$

which sets the cross terms in the variance of $g(z, V)$ to zero. From (42), (44) we have

$$\begin{aligned} var(g(z, V^N)) &= \frac{1}{N^2} \sum_{i=1}^N var(g(z, V_i)) \\ &= \frac{1}{N} F(z)(1 - F(z)) \end{aligned} \quad (45)$$

APPENDIX III

MEAN AND VARIANCE OF THE DATA WITH SORTING SIGNATURE

$$\begin{aligned} E(g(z, \theta_i)) &= Pr(|v_i + \bar{\theta}_i| \leq z) \\ &= Pr(-z - \bar{\theta}_i \leq v_i \leq z - \bar{\theta}_i) \\ &= \phi\left(\frac{z - \bar{\theta}_i}{\sigma}\right) - \phi\left(\frac{-z - \bar{\theta}_i}{\sigma}\right) \end{aligned} \quad (46)$$

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

$$-\phi\left(\frac{-z - \bar{\theta}_i}{\sigma}\right) = \phi\left(\frac{z + \bar{\theta}_i}{\sigma}\right) - 1 \quad (47)$$

which results $E(g(z, \theta_i))$ to be the defined $H(z, \bar{\theta}_i)$ in (27). For the variance, due to the structure of $g(z, \bar{\theta}_i)$, similar to what we had for the noise, we have

$$E(g^2(z, \theta_i)) = E(g(z, \theta_i)) = H(z, \bar{\theta}_i). \quad (48)$$

Therefore

$$\begin{aligned} var(g(z, \theta_i)) &= E(g^2(z, \theta_i)) - (E(g(z, \theta_i)))^2 \\ &= H(z, \bar{\theta}_i)(1 - H(z, \bar{\theta}_i)). \end{aligned} \quad (49)$$

Which results in the variance in (30).

APPENDIX IV

MEAN AND VARIANCE OF THE SIGNATURE FOR COLORED NOISE

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

$$\begin{aligned} g(z, v^N) &= \frac{1}{N} \sum_{i=1}^N g(z, v_i) \\ &= \left[\frac{1}{N} \quad \frac{1}{N} \quad \cdots \quad \frac{1}{N} \right] \begin{bmatrix} g(z, v_1) \\ g(z, v_2) \\ \vdots \\ g(z, v_N) \end{bmatrix}. \end{aligned} \quad (50)$$

For the expected value of this function we have

$$E(g(z, V^N)) = \frac{1}{N} \sum_{i=1}^N E(g(z, V_i)) = F(z) \quad (51)$$

which is similar to that of the i.i.d. additive noise. However, for the variance, since (52) holds [see (52) at the bottom of the page], we have

$$\begin{aligned} var(g(z, V^N)) &= \frac{1}{N} F(z)(1 - F(z)) \\ &\quad + \frac{1}{N^2} \sum_{i=1, j=1, i \neq j}^N cov(g(z, V_i)g(z, V_j)) \end{aligned} \quad (53)$$

where elements of the second term are

$$\begin{aligned} cov(g(z, V_i)g(z, V_j)) \\ = E(g(z, V_i)(g(z, V_j))) - E(g(z, V_i))E(g(z, V_j)) \end{aligned} \quad (54)$$

$$\begin{aligned} var(g(z, V^N)) &= var\left(\frac{1}{N} \sum_{i=1}^N g(z, V_i)\right) \\ &= \left[\frac{1}{N} \quad \frac{1}{N} \quad \cdots \quad \frac{1}{N} \right] \\ &\quad \times \begin{bmatrix} var(g(z, V_1)) & cov(g(z, V_1)g(z, V_2)) & \cdots & cov(g(z, V_1)g(z, V_N)) \\ cov(g(z, V_1)g(z, V_2)) & var(g(z, V_2)) & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ cov(g(z, V_1)g(z, V_N)) & \cdots & \cdots & var(g(z, V_N)) \end{bmatrix} \begin{bmatrix} \frac{1}{N} \\ \frac{1}{N} \\ \vdots \\ \frac{1}{N} \end{bmatrix} \end{aligned} \quad (52)$$

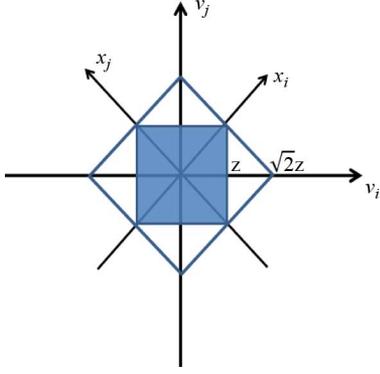


Fig. 12. The desired area for calculation of the probabilities in (59) and (61).

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

$$\begin{aligned} E \begin{pmatrix} V_i \\ V_j \end{pmatrix} &= \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \\ \text{var} \begin{pmatrix} V_i \\ V_j \end{pmatrix} &= \sigma^2 \begin{bmatrix} 1 & \rho_{ij} \\ \rho_{ij} & 1 \end{bmatrix} \end{aligned} \quad (55)$$

where

$$\rho_{ij} = \frac{R_{vv}(i-j)}{R_{vv}(0)} \quad (56)$$

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

$$\begin{bmatrix} 1 & \rho_{ij} \\ \rho_{ij} & 1 \end{bmatrix} = \frac{1}{2} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} 1+\rho_{ij} & 0 \\ 0 & 1-\rho_{ij} \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}. \quad (57)$$

Therefore, by the following transformation:

$$\begin{bmatrix} x_i \\ x_j \end{bmatrix} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} v_i \\ v_j \end{bmatrix} \quad (58)$$

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

$$\begin{aligned} E(g(z, V_i)g(z, V_j)) & \\ &= Pr(|V_i| \leq z \& |V_j| \leq z) \end{aligned} \quad (59)$$

$$\leq Pr(|X_i| \leq \sqrt{2}z \& |X_j| \leq \sqrt{2}z) \quad (60)$$

$$= F \left(\frac{\sqrt{2}z}{\sqrt{1+\rho_{ij}}} \right) F \left(\frac{\sqrt{2}z}{\sqrt{1-\rho_{ij}}} \right). \quad (61)$$

Fig. 12 show the area considered for the calculation of this probability.

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

$$E(g(z, \Theta^N)) = \frac{1}{N} \sum_{i=1}^N H(z, \bar{\theta}_i) \quad (62)$$

where $H(z, \bar{\theta}_i)$ was defined in (27). For the variance of the sorted absolute value of the noisy data, similarly (48) holds. Therefore, structure of the variance is similar to (53)

$$\begin{aligned} \text{var}(g(z, \Theta^N)) &= \frac{1}{N^2} \sum_{i=1}^N (H(z, \bar{\theta}_i) - H^2(z, \bar{\theta}_i)) \\ &+ \frac{1}{N^2} \sum_{i=1, j=1, i \neq j}^N \text{cov}(g(z, \theta_i)g(z, \theta_j)). \end{aligned} \quad (63)$$

For the covariance in the second term

$$\begin{aligned} \text{cov}(g(z, \Theta_i)g(z, \Theta_j)) & \\ &= E(g(z, \Theta_i)g(z, \Theta_j)) - E(g(z, \Theta_i))E(g(z, \Theta_j)) \end{aligned} \quad (64)$$

we use (62) to calculate $E(g(z, \Theta_i))E(g(z, \Theta_j))$ and have

$$\begin{aligned} E(g(z, \Theta_i)g(z, \Theta_j)) & \\ &= Pr(|V_i + \bar{\theta}_i| \leq z \& |V_j + \bar{\theta}_j| \leq z) \end{aligned} \quad (65)$$

$$\begin{aligned} &\leq Pr \left(\sqrt{2} \left(-z - \frac{\bar{\theta}_i + \bar{\theta}_j}{2} \right) \leq X_i \leq \sqrt{2} \left(z - \frac{\bar{\theta}_i + \bar{\theta}_j}{2} \right) \right. \\ &\quad \left. \& \sqrt{2} \left(-z - \frac{\bar{\theta}_i - \bar{\theta}_j}{2} \right) \leq X_j \leq \sqrt{2} \left(z - \frac{\bar{\theta}_i - \bar{\theta}_j}{2} \right) \right) \end{aligned} \quad (66)$$

$$\begin{aligned} &= H \left(\frac{\sqrt{2}z}{\sqrt{1+\rho_{ij}}}, \frac{\bar{\theta}_i + \bar{\theta}_j}{\sqrt{2(1+\rho_{ij})}} \right) \\ &\quad \times H \left(\frac{\sqrt{2}z}{\sqrt{1-\rho_{ij}}}, \frac{\bar{\theta}_i - \bar{\theta}_j}{\sqrt{2(1-\rho_{ij})}} \right). \end{aligned} \quad (67)$$

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," *J. Amer. Statist. Assoc.*, vol. 90, pp. 1200–1224, 1995.
- [3] T. Blu and F. Luisier, "The SURE-LET approach to image denoising," *IEEE Trans. Image Process.*, vol. 16, pp. 2778–2786, Nov. 2007.
- [4] X.-P. Zhang, "Thresholding neural network for adaptive noise reduction," *IEEE Trans. Neural Netw.*, vol. 12, pp. 567–584, May 2001.
- [5] S. G. Chang, B. Yu, and M. Vetterli, "Adaptive wavelet thresholding for image denoising and compression," *IEEE Trans. Image Process.*, vol. 9, pp. 1532–1546, Sep. 2000.
- [6] S. G. Chang, B. Yu, and M. Vetterli, "Spatially adaptive wavelet thresholding with context modeling for image denoising," *IEEE Trans. Image Process.*, vol. 9, pp. 1522–1531, Sep. 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 Trans. Image Process.*, vol. 15, pp. 654–665, Mar. 2006.
- [9] S. Beheshti, N. Nikvand, and X. N. Fernando, "Soft thresholding by noise invalidation," in *Proc. 24th Biennial Symp. Commun.*, 2009, pp. 235–238.
- [10] A. H. David and H. N. Nagaraja, *Order Statistics*, 3rd ed. Hoboken, NJ: Wiley, 2003.
- [11] L. Wasserman, *All of Nonparametric Statistics*, 1st ed. New York: Springer, 2005.
- [12] Y. Blanco, S. Zazo, and J. C. Principe, "Alternative statistical Gaussianity measure using the cumulative density function," in *Proc. Int. Workshop on ICA and Blind Signal Separation*, 2000, pp. 537–542.



Soosan Beheshti (M'03–SM'06) received the B.S. degree from Isfahan University of Technology, Isfahan, Iran, and the M.S. and Ph.D. degrees from the Massachusetts Institute of Technology (MIT), Cambridge, in 1996 and 2002, respectively, all in electrical engineering.

From September 2002 to June 2005, she was a Postdoctoral Associate and a Lecturer at MIT. Since July 2005, she has been with the Department of Electrical and Computer Engineering, Ryerson University, Toronto, ON, Canada, where she is currently an Assistant Professor and director of Signal and Information Processing (SIP) Laboratory. Her research interests include statistical signal processing, hyperspectral imaging, and system dynamics and modeling.



Masoud Hashemi (S'09) received the B.S. and M.S. degrees in 2005 and 2008, respectively, all in electrical engineering from Isfahan University of Technology, Isfahan, Iran.

He is currently pursuing the M.S. degree with the Department of Electrical and Computer Engineering, Ryerson University, Toronto, Canada. His research interests include noise removal in images and biomedical signals.



Xiao-Ping Zhang (M'95–SM'06) received the B.S. and Ph.D. degrees from Tsinghua University, in 1992 and 1996, respectively, all in electronic engineering. He received the M.B.A. degree in finance and economics with Honors from the University of Chicago Booth School of Business, Chicago, IL.

Since fall 2000, he has been with the Department of Electrical and Computer Engineering, Ryerson University, where he is now Professor and Director of Communication and Signal Processing Applications Laboratory (CASPAL). Prior to joining Ryerson, from 1996 to 1998, he was a Postdoctoral Fellow at the University of Texas, San Antonio, and then at the Beckman Institute, the University of Illinois at Urbana-Champaign. He held research and teaching positions with the Communication Research Laboratory, McMaster University, Canada, in 1999. From 1999 to 2000, he was a Senior DSP Engineer with SAM Technology, Inc., San Francisco, CA, and a consultant with the San Francisco Brain Research Institute. His research interests include signal processing for communications, multimedia retrieval and video content analysis, computational intelligence, and applications in bioinformatics, finance, and marketing. He is a frequent consultant for biotech companies and investment firms.

Dr. Zhang is a registered Professional Engineer in Ontario, Canada and a member of Beta Gamma Sigma Honor Society. He is the Publicity Co-Chair for ICME'06 and program Co-Chair for ICIC'05.



Nima Nikvand (S'10) received the B.S. degree from Isfahan University of Technology, Isfahan, Iran, and the M.S. degrees from Ryerson University, Toronto, Canada, in 2008, all in electrical engineering.

He is currently pursuing the Ph.D. degree with the Department of Electrical and Computer Engineering, University of Waterloo. His research interests include signal and image processing and information theoretic approaches in signal quality assessment.