Experiments in Wavelet Shrinkage Denoising

Carl Taswell*

Abstract

Previous simulation experiments for the comparison of wavelet shrinkage denoising methods have failed to demonstrate significant differences between methods. Such differences have never been clearly demonstrated due to the use of qualitative comparisons or of quantitative comparisons that suffered from insufficient sample size and/or absent confidence intervals for the figure of merit investigated.

In particular, previous studies have used non-robust measures as figures of merit for *fixed* signal classes defined by adding instances of noise to the same instance of the fixed test signal. New simulation experiments are reported here that instead use robust measures for *randomized* signal classes defined by adding instances of noise to different instances of randomized test signals.

Significantly greater variability in the performance of the denoising methods was observed when comparing results obtained with randomized rather than fixed signal classes. However, the use of robust measures does facilitate statistically valid comparisons with respect to this variability. Indeed, the use of non-robust or of non-randomized signal classes can result in misleading inferences from invalid comparisons. Thus, the combined use of both should yield more realistic and meaningful simulation results that better represent the real-world context intended for applied use of the denoising methods.

Keywords: wavelet-based denoising, wavelet domain thresholding, wavelet shrinkage, noise removal, non-parametric signal estimation, robust measures.

1 Introduction

Denoising should not be confused with *smoothing*. Whereas smoothing removes high frequencies and retains low frequencies, denoising attempts to remove whatever noise is present and retain whatever signal is present regardless of the spectral content of the noisy signal. For example, to denoise music corrupted by noise, the high frequencies of the music should not be eliminated. Instead, both the treble and the bass should be preserved. Although not demonstrated here, this example of denoising music offers an important application of wavelet shrinkage denoising for further investigation.

As developed originally by Donoho *et al.* [4, 5, 3, 6], wavelet shrinkage denoising is *denoising by shrinking* (*i.e.*, nonlinear soft thresholding) *coefficients in the wavelet transform domain*. It consists of three steps: 1) a linear forward wavelet transform, 2) a nonlinear shrinkage denoising, and 3) a linear inverse wavelet transform. Because of the *nonlinear* shrinking of coefficients in the transform domain, this procedure is distinct from those denoising methods that are

*C. Taswell (ctaswell@toolsmiths.com) is with Computational Toolsmiths, POB 18925, Stanford, CA 94309-8925.

entirely linear. Moreover, it is considered a nonparametric method. Thus, it is distinct from parametric methods [10], including both linear and nonlinear regression [7], in which parameters must be estimated for a particular model that must be assumed *a priori*. (For example, the most commonly used parametric method is least squares regression to estimate the parameters a and b in the model $y = ax + b$.)

The first Monte Carlo simulation experiment comparing any of the various wavelet shrinkage denoising procedures was performed by Taswell and published in the article by Donoho and Johnstone [4, Table 4, page 448; Acknowledgements, page 450]. Various other experiments have since been performed by other authors (see discussion and references in [6], also [8, 2]). Most of this work has examined a few well known images ('Barbara', 'Lena','baboon', *et c.*) or the four test signals originally called 'Doppler', 'HeaviSine', 'Blocks', and 'Bumps' by Donoho and Johnstone [4]. The latter was renamed more descriptively as 'Spires' by Taswell [15]. All of the experiments on these test signals, including the most recent experiments on signals [15, 8] and on images [2], examined only fixed test signals and images rather than defined classes of randomized test signals and images (or alternatively, classes of real-world signals and images with many different instances in each class).

To address this deficiency in the design of the simulation experiments, new classes of randomized test signals are introduced here, and used in new experiments which provide a more appropriate evaluation of the performance of the denoising methods. For example, instead of using just one instance of 'Spires' with the particular values of the peak height, width, and location parameters originally defined in [4], multiple instances of 'Random Spires' are used in the experiments with randomized values of the peak height, width, and location parameters. The use of such randomized signal classes in the simulation experiments results in a more realistic assessment of the variability of performance that can be expected for the different denoising methods. Some of the results presented here have previously appeared in the conference paper [13].

2 Methods

2.1 Wavelet Shrinkage Denoising

Assume that the observed data

$$X[n] = S[n] + G[n]$$

contains the true signal $S[n]$ with additive Gaussian noise $G[n]$ as functions in time at sample points n. Let $\mathcal{W}(\cdot)$ and $\mathcal{W}^{-1}(\cdot)$ denote the forward and inverse wavelet transform operators. Let $\mathcal{D}(\cdot, \lambda)$ denote the denoising operator with soft threshold λ. We intend to wavelet shrinkage denoise $X[n]$ in order to recover $\hat{S}[n]$ as an estimate of $S[n]$. Then the three steps

$$\begin{aligned} Y &= \mathcal{W}(X) \\ Z &= \mathcal{D}(Y, \lambda) \\ \hat{S} &= \mathcal{W}^{-1}(Z) \end{aligned}$$

summarize the procedure.

Given threshold λ for data U (in any arbitrary domain – signal, transform, or otherwise), the rule

$$\mathcal{D}(U, \lambda) \equiv \operatorname{sgn}(U) \max(0, |U| - \lambda)$$

defines nonlinear soft thresholding. The operator $\mathcal{D}$ nulls all values of U for which $|U| \leq \lambda$ and shrinks toward the origin by an amount λ all values of U for which $|U| > \lambda$. It is the latter aspect that has led to $\mathcal{D}$ being called the shrinkage operator in addition to the soft thresholding operator.

To determine λ, let's say that the data has sample size or length N if it has been sampled at N time points n_i such that $X_i \equiv X[n_i]$. Then for an orthogonal $\mathcal{W}$, there will also be N transform coefficients Y_j. If we prefer to use a threshold (such as the minimax threshold or the universal threshold [4]) that depends only on N, then λ can be predetermined and we can use the three-step denoising procedure already described. However, if we prefer to use a data-adaptive threshold

$$\lambda = d(U)$$

(such as the threshold selected by Stein's unbiased risk estimator (SURE) [5]) that depends not just on N but on U (which again represents the data in any generic domain), then we must use a four-step procedure

$$\begin{aligned} Y &= \mathcal{W}(X) \\ \lambda &= d(Y) \\ Z &= \mathcal{D}(Y, \lambda) \\ \hat{S} &= \mathcal{W}^{-1}(Z) \end{aligned}$$

for wavelet shrinkage denoising. Note the distinction between the operator $d(\cdot)$ that selects the threshold and the operator $\mathcal{D}(\cdot, \cdot)$ that performs the thresholding.

Implementation of $\mathcal{W}$ will not be reviewed here. Recall, however, that a wavelet transform must be specified by its analysis and synthesis wavelet filter banks, single-level convolutions and boundary treatment, and the total number L of iterated multiresolution levels [12]. Thus, we can generate many different kinds of wavelet shrinkage denoising procedures by combining different choices for $\mathcal{W}(\cdot)$ and $d(\cdot)$. If we let $\mathcal{D}$ denote more generally either the soft thresholding operator $\mathcal{D}_\mathrm{s}$ or the hard thresholding operator $\mathcal{D}_\mathrm{h}$ [4], then by combining choices for $\mathcal{W}(\cdot)$, $\mathcal{D}(\cdot, \cdot)$, and $d(\cdot)$, we can generate even more different kinds of wavelet-based denoising.

Denoising by thresholding in the wavelet domain has been developed principally by Donoho *et al.* [4, 5, 3, 6]. In [4], they introduced *RiskShrink* with the minimax threshold, *VisuShrink* with the universal threshold, and discussed both hard and soft thresholds in a general context that included ideal denoising in both the wavelet and Fourier domains. In [5], they introduced *SureShrink* with the SURE threshold, *WaveJS* with the James-Stein threshold, and *LPJS* also with the James-Stein threshold but in the Fourier domain instead of the wavelet domain. The procedure *LPJS* was renamed *FourJS* (analogous to *WaveJS*) for consistency of mnemonics by Taswell [15], who also labelled the various denoising procedures respectively 'RIS', 'VIS', 'IWD', 'IFD', 'SUR', 'WJS', and 'FJS' for use as abbreviations.

These procedures can be classified by transform domain, Fourier versus wavelet, as well as by intent of use, ideal versus practical. An ideal procedure requires *a priori* knowledge of the noise, whereas a practical procedure does not, so that ideal procedures are only used for purposes of comparison in simulation experiments. Moreover, the procedures can be classified according to whether they use a single threshold globally for all relevant parts of the transform, or multiple thresholds locally for different parts of the transform (Fourier frequency bands or wavelet multiresolution levels). For example, 'VisuShrink' ('VIS') is a practical, wavelet domain, global threshold procedure in which $\lambda = \sqrt{2 \log N}$ is used for all levels $l = 1, \ldots, L$ from fine to coarse. As another example, 'SureShrink' ('SUR') is also a practical wavelet procedure but it uses a local threshold λ_l estimated adaptively for each level l.

Wavelet shrinkage denoising results reported here were generated with Version 4.6c1 of the WAVBOX Software Library [11] using orthogonal discrete wavelet transforms and wavelet filters from the systematized collection of Daubechies wavelets [14], in particular, DROLA(10;5).

2.2 Simulation Experiments

Randomized signal classes, called 'Random Blocks', 'Random Spires', and 'Random HeaviSine', were defined to generate signals analogous to the original 'Blocks', 'Spires', and 'HeaviSine'. Figure 1 displays the original nonrandomized versions in the top row of subplots, and one instance each of the randomized versions in the bottom row of subplots. Table 1 lists the mathematical formulae for the test signal classes. These formulae are valid for both the randomized and original nonrandomized versions with the appropriate choice of parameters. Table 2 lists the MATLAB pseudocode expressions for the set of parameters chosen for the randomized classes used in the experiments reported here.

Instances of randomized signals were generated from the signal classes, corrupted with additive Gaussian noise with a signal-to-noise ratio SNR = 10, and then denoised with each of the various denoising methods. Performance of the denoising procedures on the signal classes was studied as a function of signal length $N = 2^J$ over T trials of the simulation. Several criteria, including the SNR and the ℓ^1, ℓ^2, and ℓ^∞ norms, were used as objective figures of merit for comparing the original signal $S_t[n]$ with the denoised estimate $\hat{S}_t[n]$ in the t^{th} trial. Results were averaged over signal instances for all trials $t = 1, \ldots, T$ in each signal class and reported with means, standard deviations, and coefficients of variation. SNR was computed as

$$\text{SNR}(S_t, \hat{S}_t) = 10 \log_{10} \frac{\sum_{n=1}^{N} |S_t[n]|^2}{\sum_{n=1}^{N} |S_t[n] - \hat{S}_t[n]|^2}$$

in decibels (dB). For each figure of merit in each trial, the rank order of the various denoising procedures as determined by that figure of merit was computed and then averaged over all trials. Table 3 summarizes the basic design parameters for the Monte Carlo simulation experiments used for all signal classes and denoising procedures.

3 Results

As expected, the coefficients of variation (ratios of standard deviation to mean) were observed to be significantly larger for the randomized signal classes when compared with the original fixed signal classes. Table 4 displays an example with 'VIS' that demonstrates an increase of approximately 2 – 10 fold for SNR values. Analogous results with increased coefficients of variation for SNR values were also observed for the other methods with the amount of increase dependent upon signal class, signal length, and denoising procedure. However, Table 5 displays an example with 'SUR' that demonstrates greater robustness, *i.e.*,, less difference between fixed and random signal classes, when SNR ranks instead of SNR values were used in the comparison. Nevertheless, in most cases, the random signal classes manifested greater variability than the fixed signal classes.

Continuing the 'SUR' example, Tables 6 and 7 present the SNR values and ranks with results expressed as the mean ± 1 standard deviation from all trials. Recall that SNR values are reported in dB while SNR ranks are reported in the interval $[1, 7]$ because there were seven different denoising procedures tested and then ranked with 1 assigned to the worst (with the lowest SNR dB value) and 7 to the best (with the highest SNR dB value). Note again the greater differences between fixed and random classes with the use of SNR values versus ranks.

Figures 2 and 3 plots curves with results, respectively, for the SNR values and ranks for all seven denoising procedures as functions of $J = \log_2 N$ with each point and error bar corresponding to the mean ± 1 standard deviation from all trials. Large differences between fixed and random classes necessarily impact the statistical validity of comparisons of the denoising procedures. For example, in Figure 2 when comparing performance by SNR values of the methods on the signal classes, the error bars do not overlap for 'Blocks' but do overlap for 'Random Blocks', implying that any differences between the methods are not statistically significant for this randomized signal class under the experimental conditions investigated. However, in Figure 3 when comparing performance by SNR ranks, the error bars for 'VIS' and 'SUR' do not overlap for 'Random Blocks', implying that in this case the method 'VIS' does perform significantly worse than the method 'SUR'.

Assuming that the randomized classes are a more appropriate simulation of real-world situations and data, then the following results can be summarized from inspection of the bottom row subplots in Figure 3: As expected by theory, ideal wavelet denoising ('IWD') performed best. Of the non-ideal or practical procedures investigated, VisuShrink ('VIS') and SureShrink ('SUR') per-

formed, respectively, worst and best in the majority of cases studied (*i.e.*,for the various signal classes and lengths).

4 Discussion

Monte Carlo simulation experiments have been performed previously [4, 6, 1, 15, 9, 16, 2, 8] in an effort to compare various wavelet-based denoising procedures. Typically, these experiments have involved taking real signals or generating synthetic signals, adding known amounts of noise, and then comparing the effectiveness with which various denoising procedures remove the noise. Unfortunately, the statistical significance of any differences between the various methods has never been clearly demonstrated due to the use of qualitative comparisons by human subjective opinion or of quantitative comparisons that suffered from insufficient sample size (*i.e.*, limited number of distinct test signals in the defined class) and/or absent confidence intervals (*i.e.*, without implicit or explicit hypothesis testing) for the figure of merit investigated.

In one of the few studies that reported actual statistical test results, Wachowiak *et al.* [16] clearly stated: "Paired Student's t-tests failed to show significant differences among the denoising techniques with respect to any merit measure... performance was quantified on the basis of the averaged merit measure." However, the results reported here in Section 3 with the new randomized signal classes defined in Section 2.2 demonstrate that the use of averaged merit measures may be inadequate to identify differences between denoising procedures. On the contrary, the use of more robust measures, such as the figure of merit's rank instead of the figure of merit itself, can more definitively establish significant differences as demonstrated by the results displayed in Figure 3 for the SNR ranks versus those in Figure 2 for the SNR values.

The use of both robust measures and randomized signal classes have been introduced here to facilitate the statistically valid comparison of the performance of wavelet shrinkage denoising methods. Use of non-robust measures or of non-randomized signal classes can result in misleading inferences from invalid comparisons. Careful attention should be focused on the use of appropriately defined measures and signal classes when evaluating denoising methods in simulation experiments. These experiments can then be used to compare the performance of various denoising methods assuming, of course, that the design parameters for both the randomized signal class and for the simulation experiment appropriately represent both the situation and the data in the real-world context intended for applied use of the denoising methods.

When there is no statistically significant difference in the methods' performance on the defined signal class for the specified experimental conditions, other criteria such as computational complexity should be used to select a preferred method. Moreover, if a particular method can be demonstrated to perform significantly worse than other competing methods, such as shown here for VisuShrink, it would be prudent to exclude it from further consideration for use as a denoising method for the signal class and experimental conditions investigated.

References

[1] Philippe Carré, Hélene Leman, Christine Fernandez, and Catherine Marque. Denoising of the uterine EHG by an undecimated wavelet transform. *IEEE Transactions on Biomedical Engineering*, 45(9):1104–1113, September 1998.

[2] S. G. Chang, Bin Yu, and Martin Vetterli. Adaptive wavelet thresholding for image denoising and compression. *IEEE Transactions on Image Processing*, 9(9):1532–1546, September 2000.

[3] David L. Donoho. Denoising via soft thresholding. *IEEE Transactions on Information Theory*, 41:613–627, May 1995.

[4] David L. Donoho and Iain M. Johnstone. Ideal spatial adaption via wavelet shrinkage. *Biometrika*, 81:425–455, September 1994.

[5] David L. Donoho and Iain M. Johnstone. Adapting to unknown smoothness via wavelet shrinkage. *Journal of the American Statistical Association*, 90(432):1200–1224, December 1995.

[6] David L. Donoho, Iain M. Johnstone, Gerard Kerkyacharian, and Dominique Picard. Wavelet shrinkage: Asymptopia? with discussion. *Journal of the Royal Statistical Society, Series B*, 57(2):301–369, 1995.

[7] A. Ronald Gallant. *Nonlinear Statistical Models*. John Wiley & Sons, Inc., New York, NY, 1987.

[8] V. P. Melnik, I. Shmulevich, K. Egiazarian, and J. Astola. Block-median pyramidal transform: Analysis and denoising applications. *IEEE Transactions on Signal Processing*, 49(2):364–372, February 2001.

[9] Quan Pan, Lei Zhang, Guanzhong Dai, and Hongcai Zhang. Two denoising methods by wavelet transform. *IEEE Transactions on Signal Processing*, 47(12):3401–3406, December 1999.

[10] Sam Shearman. Parametric curve-fitting software satisfies many needs. *Personal Engineering & Instrumentation News*, 15(9):24–31, September 1998.

[11] Carl Taswell. *WAVBOX Software Library and Reference Manual*. Computational Toolsmiths, `www.wavbox.com`, 1993–2001.

[12] Carl Taswell. Specifications and standards for reproducibility of wavelet transforms. In *Proceedings of the International Conference on Signal Processing Applications and Technology*, pages 1923–1927. Miller Freeman, October 1996.

[13] Carl Taswell. Randomized signal classes for evaluating the performance of wavelet shrinkage denoising methods. In *Proceedings of the IASTED International Conference on Signal and Image Processing*, pages 352–355, October 1999. Paper #296-214.

[14] Carl Taswell. Constraint-selected and search-optimized families of Daubechies wavelet filters computable by spectral factorization. *Journal of Computational and Applied Mathematics*, 121:179–195, 2000. Invited Paper, Special Issue on Numerical Analysis in the Twentieth Century.

[15] Carl Taswell. The what, how, and why of wavelet shrinkage denoising. *IEEE Computing in Science & Engineering*, 2(3):12–19, May–June 2000.

[16] Mark P. Wachowiak, Gregory S. Rash, Peter M. Quesada, and Ahmed H. Desoky. Wavelet-based noise removal for biomechanical signals: A comparative study. *IEEE Transactions on Biomedical Engineering*, 47(3):360–368, March 2000.

Table 1: Mathematical Formulae for Signal Classes

Name	Function	Kernel	Parameters
Blocks	$f(t) = \sum_{m=1}^{M} h_m K(t - p_m)$	$K(s) = (\mathrm{sgn}(s) + 1)/2$	M, h_m, p_m
Spires	$f(t) = \sum_{m=1}^{M} h_m K((t - p_m)/w_m)$	$K(s) = (\lvert s\rvert + 1)^{-4}$	M, h_m, p_m, w_m
HeaviSine	$f(t) = h_1 \sin(p_1 \pi t) + \sum_{m=2}^{M} h_m K(t - p_m)$	$K(s) = \mathrm{sgn}(s)$	M, h_m, p_m

Table 2: Pseudocode Expressions for Parameters Used in Randomized Versions

Name	h_m	p_m	w_m	m
Blocks	5*sign(rand(1,11)-0.5).*rand(1,11)	sort(rand(1,11))		1, . . . , 11
Spires	5*rand(1,11)	sort(rand(1,11))	0.05*rand(1,11)	1, . . . , 11
HeaviSine	4	4		1
	2*sign(rand(1,2)-0.5).*rand(1,2)	sort(rand(1,2))		2,3

Table 3: Simulation Experiment Design Parameters

T	N	J	L
100	256	8	3
100	512	9	4
100	1024	10	5
80	2048	11	6
40	4096	12	7
20	8192	13	8
10	16384	14	9

Table 4: VIS Denoising: Coefficients of Variation for SNR Values.

	Blocks		Spires		HeaviSine	
J	Fixed	Random	Fixed	Random	Fixed	Random
8	2.8501e-002	1.1718e-001	7.9789e-002	1.0950e-001	5.2748e-002	6.5265e-002
9	3.4988e-002	1.3192e-001	5.5483e-002	8.5289e-002	3.5886e-002	5.7628e-002
10	2.3885e-002	1.3916e-001	3.5810e-002	7.8460e-002	3.0110e-002	6.2710e-002
11	1.8842e-002	1.4474e-001	1.9456e-002	5.7327e-002	2.3419e-002	6.8886e-002
12	1.5847e-002	1.2818e-001	1.6622e-002	5.0351e-002	2.4020e-002	5.3679e-002
13	1.0527e-002	8.9286e-002	1.0516e-002	4.1403e-002	1.2165e-002	7.0534e-002
14	6.7975e-003	1.1607e-001	7.8823e-003	5.1193e-002	1.4964e-002	1.1089e-001

Table 5: SUR Denoising: Coefficients of Variation for SNR Ranks.

	Blocks		Spires		HeaviSine	
J	Fixed	Random	Fixed	Random	Fixed	Random
8	2.4330e-001	3.1641e-001	3.8317e-001	4.1271e-001	2.3139e-001	2.1446e-001
9	2.8785e-001	3.4927e-001	8.9728e-002	2.3793e-001	2.3570e-001	2.4149e-001
10	2.1884e-001	3.5904e-001	9.1179e-002	1.9156e-001	2.2765e-001	2.1524e-001
11	2.4530e-001	2.8633e-001	2.6295e-002	1.4172e-001	1.9139e-001	1.8090e-001
12	8.4628e-002	2.2624e-001	0	1.1720e-001	1.3020e-001	1.7733e-001
13	9.3271e-002	2.0194e-001	3.7581e-002	7.0757e-002	3.7581e-002	6.2624e-002
14	0	5.3598e-002	0	0	0	0

Table 6: SUR Denoising: SNR Values in dB.

	Blocks		Spires		HeaviSine	
J	Fixed	Random	Fixed	Random	Fixed	Random
8	11.8±0.74	14.8±1.54	10.7±1.34	12.3±0.91	18.4±0.99	18.1±0.96
9	13.9±0.74	16.9±1.72	12.8±0.47	14.6±0.99	20.5±0.76	20.5±1.13
10	15.7±0.33	18.7±1.98	14.2±0.51	16.8±0.96	22.7±0.75	22.2±1.38
11	17.1±0.44	20.8±1.89	16.7±0.27	19.1±0.86	24.7±0.69	24.2±1.49
12	19.1±0.26	22.9±1.76	19.0±0.23	21.4±0.82	27.0±0.85	26.3±1.42
13	20.9±0.47	24.8±1.44	21.6±0.54	23.8±0.84	28.9±0.44	28.4±1.77
14	23.1±0.24	26.9±0.90	24.1±0.19	26.5±0.97	30.6±0.57	31.4±1.73

Table 7: SUR Denoising: SNR Ranks $\in [1, 7]$.

	Blocks		Spires		HeaviSine	
J	Fixed	Random	Fixed	Random	Fixed	Random
8	2.30±0.56	2.48±0.78	3.94±1.51	3.39±1.40	3.98±0.92	4.14±0.89
9	3.27±0.94	2.69±0.94	5.56±0.50	4.49±1.07	4.00±0.94	4.46±1.08
10	3.61±0.79	3.16±1.13	5.40±0.49	4.71±0.90	4.24±0.97	4.84±1.04
11	3.83±0.94	4.08±1.17	5.98±0.16	5.50±0.78	4.48±0.86	4.91±0.89
12	4.93±0.42	4.48±1.01	6.00±0.00	5.65±0.66	5.50±0.72	5.33±0.94
13	5.50±0.51	5.15±1.04	5.95±0.22	5.80±0.41	5.95±0.22	5.85±0.37
14	6.00±0.00	5.90±0.32	6.00±0.00	6.00±0.00	6.00±0.00	6.00±0.00

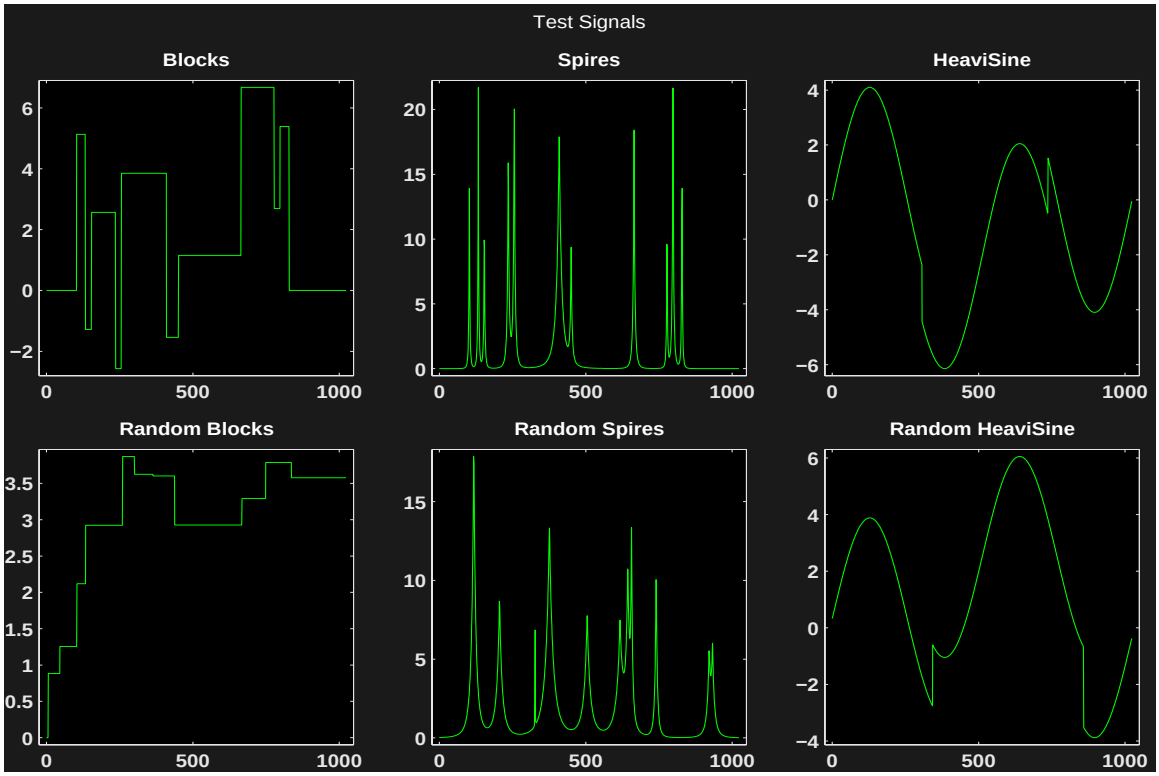

Figure 1: Standardized Test Signals.

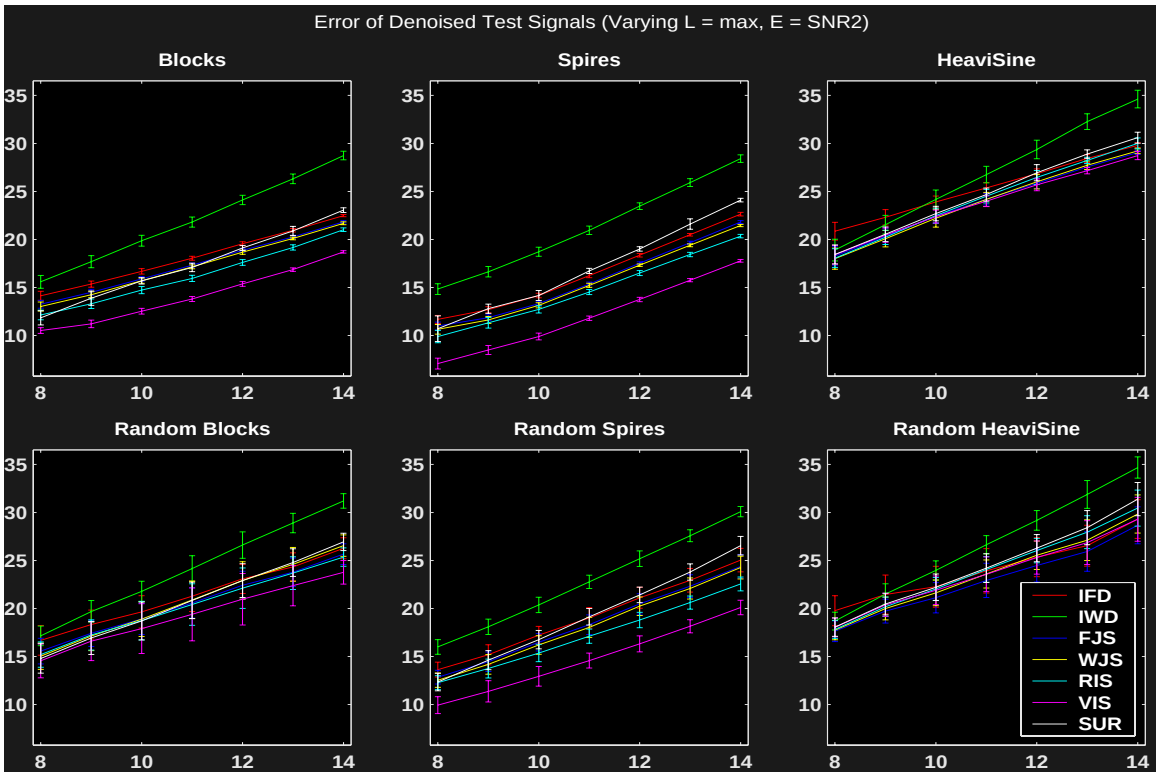

Figure 2: SNR Values in dB for Denoised Test Signals.

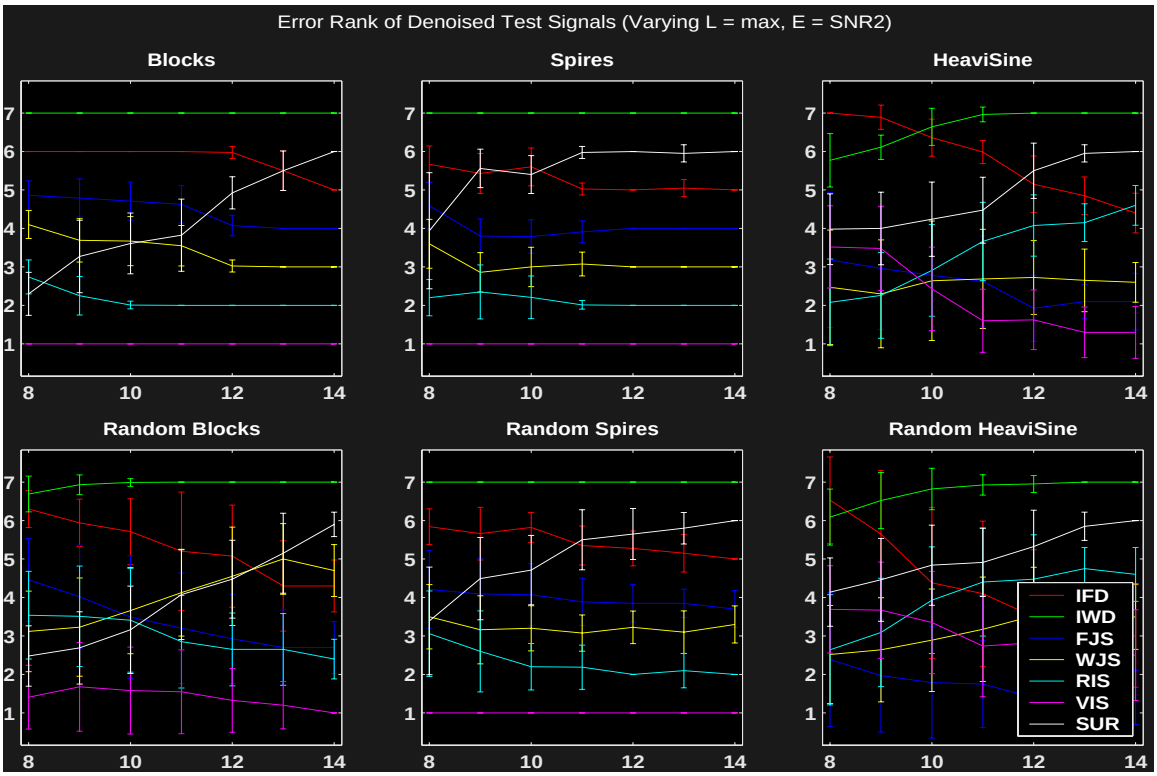

Figure 3: SNR Ranks $\in [1, 7]$ for Denoised Test Signals.