Improved Wavelet Denoising via Empirical Wiener Filtering

Sandeep P. Ghael, Akbar M. Sayeed, and Richard G. Baraniuk

Department of Electrical and Computer Engineering
Rice University
Houston, TX 77005 USA

Proceedings of SPIE, Mathematical Imaging, San Diego, July 1997.

ABSTRACT

Wavelet shrinkage is a signal estimation technique that exploits the remarkable abilities of the wavelet transform for signal compression. Wavelet shrinkage using thresholding is asymptotically optimal in a minimax mean-square error (MSE) sense over a variety of smoothness spaces. However, for any given signal, the MSE-optimal processing is achieved by the Wiener filter, which delivers substantially improved performance. In this paper, we develop a new algorithm for wavelet denoising that uses a wavelet shrinkage estimate as a means to design a wavelet-domain Wiener filter. The shrinkage estimate indirectly yields an estimate of the signal subspace that is leveraged into the design of the filter. A peculiar aspect of the algorithm is its use of two wavelet bases: one for the design of the empirical Wiener filter and one for its application. Simulation results show up to a factor of 2 improvement in MSE over wavelet shrinkage, with a corresponding improvement in visual quality of the estimate. Simulations also yield a remarkable observation: whereas shrinkage estimates typically improve performance by trading bias for variance or vice versa, the proposed scheme typically decreases both bias and variance compared to wavelet shrinkage.

Keywords: Wavelets, denoising, estimation, Wiener filter, subspace

1. INTRODUCTION

The wavelet transform has rapidly become an indispensable signal and image processing tool for a variety of applications, including estimation, classification, and compression. One of the key properties underlying the success of wavelets is that they form *unconditional bases* for a wide variety of signal classes.[1] Consequently, wavelet expansions tend to concentrate the signal energy into a relatively small number of large coefficients. This energy compaction property of the wavelet transform makes the wavelet domain attractive for signal processing, and in particular signal estimation.

Consider the classical problem of recovering samples of an unknown deterministic continuous-time signal $s(t),\ t \in (0,1]$, from the set of noise-corrupted samples

$$x(i) \ \stackrel{def}{=}\ s(i/N) \ + \ n(i), \qquad i = 1, 2, \ldots, N, \tag{1}$$

with $n(i)$ a zero-mean white Gaussian noise of variance σ^2. Let $\boldsymbol{x}$, $\boldsymbol{s}$, $\boldsymbol{n}$ denote $N \times 1$ column vectors containing the samples $x(i)$, $s(i/N)$, and $n(i)$, respectively, and let $\boldsymbol{W}$ denote an $N \times N$ orthonormal wavelet transform matrix.[2,3] In the wavelet domain, (1) becomes

$$\boldsymbol{y} \ \stackrel{def}{=}\ \boldsymbol{\theta} \ + \ \boldsymbol{z}, \tag{2}$$

with $\boldsymbol{y} = \boldsymbol{W}\boldsymbol{x}$, $\boldsymbol{\theta} = \boldsymbol{W}\boldsymbol{s}$, and $\boldsymbol{z} = \boldsymbol{W}\boldsymbol{n}$. Our goal is to estimate the true signal wavelet coefficients $\boldsymbol{\theta}$ given the noisy observations $\boldsymbol{y}$. Note that an orthonormal wavelet transformation will map $\boldsymbol{n}$ to a $\boldsymbol{z}$ that is likewise zero-mean white Gaussian with variance σ^2, while compacting typical signals $\boldsymbol{s}$ into a small number of large wavelet coefficients in $\boldsymbol{\theta}$. Thus a reasonable approach to wavelet-based signal estimation is to shrink the small entries of $\boldsymbol{y}$ (where the signal is not) while retaining the large entries of $\boldsymbol{y}$ (where the signal is). The motivation for processing the coefficients individually stems from the fact that the wavelet transform tends to decorrelate the data. That is, roughly speaking, the wavelet transform approximates the Karhunen-Loeve (KL) transform.

Other author information: Email: akira@rice.edu, akbar@rice.edu, richb@rice.edu. Web: http://www-dsp.rice.edu. This work was supported by the National Science Foundation, grant no. MIP–9457438, and the Office of Naval Research, grant no. N00014–95–1–0849.

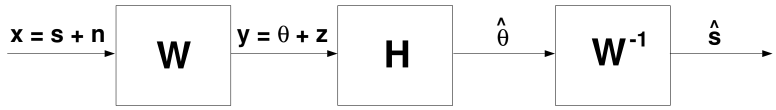

Figure 1. Wavelet-domain filtering using a diagonal weighting matrix $\boldsymbol{H}$.

While simple, the resulting *wavelet shrinkage* estimates[4,5] are surprisingly powerful. In wavelet shrinkage, the shrink/retain operation can be viewed as a diagonal filtering operation in the wavelet domain. Representing the filter by

$$\boldsymbol{H} \stackrel{def}{=} \operatorname{diag}[h(1), h(2), \ldots, h(N)], \tag{3}$$

we have the signal estimate

$$\widehat{\boldsymbol{s}} \stackrel{def}{=} \boldsymbol{W}^{-1}\boldsymbol{H}\,\boldsymbol{W}\,\boldsymbol{x}. \tag{4}$$

A block diagram for wavelet-domain filtering is given in Figure 1.

Wavelet shrinkage filters come in two basic flavors. The *hard threshold* filter $\boldsymbol{H}_h$ discards coefficients below a threshold value τ that is determined by the noise variance σ:

$$h_h(i) \stackrel{def}{=} \begin{cases} 1, & \text{if } |y(i)| > \tau \\ 0, & \text{otherwise.} \end{cases} \tag{5}$$

The *soft threshold* filter $\boldsymbol{H}_s$ is similar, but in addition shrinks the wavelet coefficients above the threshold:

$$h_s(i) \stackrel{def}{=} \begin{cases} \frac{\operatorname{sgn}[y(i)]\,[|y(i)|-\tau]}{y(i)}, & \text{if } |y(i)| > \tau \\ 0, & \text{otherwise.} \end{cases} \tag{6}$$

The uncanny ability of wavelet shrinkage to remove noise is illustrated in Figure 2 on Donoho's Doppler test signal. Related shrinkage approaches include *VisuShrink*, *SureShrink*, and the *Hybrid* scheme.[4] Certain wavelet Shrinkage estimates are asymptotically optimal in a minimax mean-squared-error (MSE) sense over various signal smoothness spaces.[4,5]

Nevertheless, for a given, finite-length signal, the optimal filter minimizing the MSE is the *Wiener filter* $\boldsymbol{H}_w$, with[6]

$$h_w(i) \stackrel{def}{=} \frac{\theta^2(i)}{\theta^2(i) + \sigma^2}. \tag{7}$$

Note that the Wiener filter requires knowledge of the signal and noise statistics. The Wiener estimate for the signal in Figure 2(a) is shown in Figure 2(g). The MSE resulting from this optimal $\boldsymbol{H}_w$ is about 3.75 times smaller than that obtained using $\boldsymbol{H}_h$.

The superior performance of the Wiener filter is due to the fact that it strikes an optimal balance in the bias-variance tradeoff. Of course, it does so only because of exact knowledge of the signal and noise statistics. In contrast, soft threshold estimates have a remarkably low variance but a relatively high bias. Other shrinkage techniques with different bias-variance tradeoffs improve the MSE by trading a small increase in variance for a substantial reduction in bias and vice versa (see[7] and the references therein.)

In this paper, we introduce a simple scheme for wavelet denoising that employs wavelet shrinkage in the design of an empirical wavelet-domain Wiener filter. Our *WienerShrink* approach involves the computation of two different wavelet transforms — one for the design of the Wiener filter and the other for its application on the noisy data. Simulation results show that *WienerShrink* signal estimates can reduce the MSE by up to a factor of 2 over classical wavelet shrinkage techniques, with a corresponding improvement in visual appearance. A *WienerShrink* denoised version of the corrupted Doppler signal from Figure 2(b) is shown in Figure 2(f). Our simulations also indicate a remarkable property of the proposed method: rather than merely striking a better bias-variance tradeoff than wavelet shrinkage, it typically improves *both* bias and variance performance simultaneously.

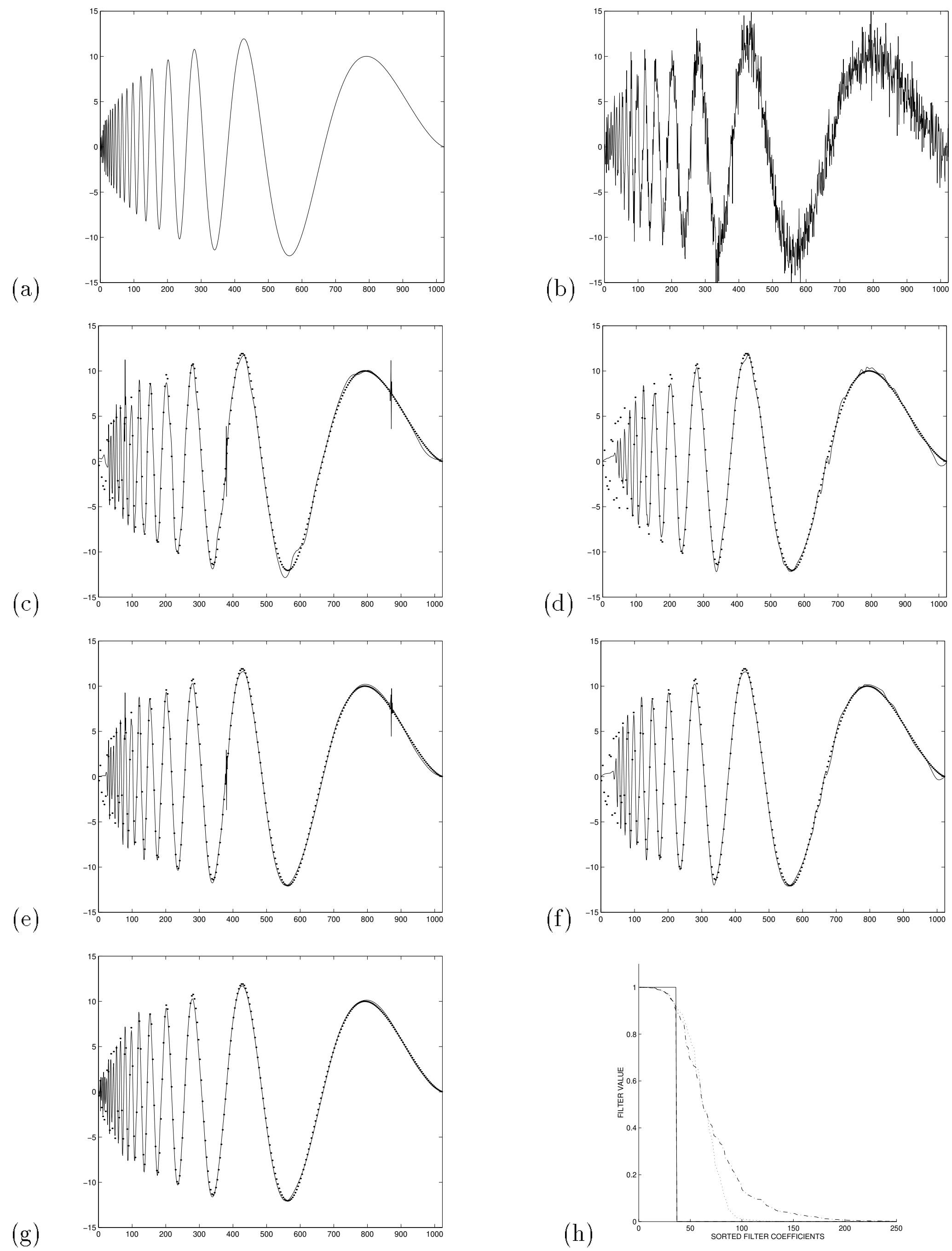

Figure 2. Signal denoising of the Doppler test signal. (a) Clean signal $\boldsymbol{s}$, $N = 1024$. (b) Noisy signal $\boldsymbol{s} + \boldsymbol{n}$ with $\sigma = \sqrt{3}$. Signal estimates using (c) hard threshold (MSE=0.58), (d) *HybridShrink* (MSE=0.71), (e) empirical Wiener filter based on hard threshold pilot estimate (MSE=0.34), (f) empirical Wiener filter based on *HybridShrink* pilot estimate (MSE=0.44), and (g) optimal Wiener filter (MSE=0.15). In each plot, the clean signal is shown dotted. (h) Wavelet-domain filters: hard threshold (solid), optimal Wiener (dot), and empirical Wiener designed using a hard threshold pilot estimate in Figure 3 with $\boldsymbol{W}_2 \neq \boldsymbol{W}_1$ (dash-dot). The empirical Wiener filter designed using $\boldsymbol{W}_2 = \boldsymbol{W}_1$ (dash) almost completely coincides with the hard threshold filter (solid).

In the next Section, we provide a discussion of the relevant aspects of the Wiener filter theory. Intuition behind *WienerShrink* follows in Section 3. In Section 4, we provide a number of examples of the method for denoising both signals and images. We close with some concluding remarks and avenues for future research in Section 5. (For background information on wavelet transforms and wavelet matrices, we refer the reader to.[2,3])

2. WIENER FILTERING

Consider anew the problem of estimating s from x in (1). The optimal linear estimator $\widehat{\boldsymbol{s}} = \boldsymbol{G}\boldsymbol{x}$ that minimizes the MSE $\mathrm{E}\|\boldsymbol{s} - \widehat{\boldsymbol{s}}\|^2$ is the Wiener filter[6]

$$\boldsymbol{G}_w \stackrel{def}{=} \boldsymbol{R}_s(\boldsymbol{R}_s + \sigma^2 \boldsymbol{I})^{-1}, \tag{8}$$

with $\boldsymbol{R}_s \stackrel{def}{=} \mathrm{E}[\boldsymbol{s}\boldsymbol{s}^T]$ and $\boldsymbol{I}$ the identity matrix.

The eigenexpansion of $\boldsymbol{R}_s$ is given by

$$\boldsymbol{R}_s \;=\; \boldsymbol{U}\Lambda\boldsymbol{U}^T \;=\; \sum_{k=1}^{N_s} \lambda_k \boldsymbol{u}_k \boldsymbol{u}_k^T. \tag{9}$$

Here, $N_s \leq N$ is the rank of $\boldsymbol{R}_s$, the $\boldsymbol{u}_k$'s are the eigenfunctions of $\boldsymbol{R}_s$, and the λ_k's are eigenvalues of $\boldsymbol{R}_s$. Stack the eigenvectors into the $N \times N_s$ unitary matrix $\boldsymbol{U} \stackrel{def}{=} [\boldsymbol{u}_1, \boldsymbol{u}_2, \ldots, \boldsymbol{u}_{N_s}]$, and place the eigenvalues into the diagonal matrix $\boldsymbol{\Lambda} \stackrel{def}{=} \mathrm{diag}[\lambda_1, \lambda_2, \ldots, \lambda_{N_s}]$. The matrix $\boldsymbol{U}$ defines the KL transform for the signal $\boldsymbol{s}$ — it decorrelates the correlation function $\boldsymbol{R}_s$ and concentrates the signal energy into the smallest possible subspace. The range space of $\boldsymbol{U}$ is the N_s-dimensional *signal subspace* $\mathcal{S}$ in which the signal energy is confined.

In terms of the eigenexpansion, the Wiener filter (8) becomes

$$\boldsymbol{G}_w \;=\; \boldsymbol{U}\widetilde{\boldsymbol{\Lambda}}\,\boldsymbol{U}^T \;=\; \sum_{k=1}^{N_s} \frac{\lambda_k}{\lambda_k + \sigma^2}\, \boldsymbol{u}_k \boldsymbol{u}_k^T, \tag{10}$$

with

$$\widetilde{\boldsymbol{\Lambda}} \stackrel{def}{=} \mathrm{diag}\left[\frac{\lambda_1}{\lambda_1 + \sigma^2}, \frac{\lambda_2}{\lambda_2 + \sigma^2}, \ldots, \frac{\lambda_{N_s}}{\lambda_{N_s} + \sigma^2}\right]. \tag{11}$$

The formulation (10) shows that the Wiener estimate is confined to the signal subspace $\mathcal{S}$ (the noise in the orthogonal complement of $\mathcal{S}$ is zeroed out). Moreover, in the eigen (KL) domain, the Wiener filter takes on the diagonal form $\widetilde{\boldsymbol{\Lambda}}$. In words, the Wiener filter first KL transforms the data via $\boldsymbol{U}^T$, processes each coefficient individually via the MMSE-optimal weighting $\widetilde{\boldsymbol{\Lambda}}$, and then transforms the data back to the original domain via $\boldsymbol{U}$.

Since the wavelet transform approximately decorrelates and concentrates the signal energy in a relatively small subspace, it can serve as an approximate KL basis for a broad class of signals. This justifies the diagonal processing of Figure 1. Wavelet shrinkage also implicitly yields an estimate of the signal subspace $\mathcal{S}$ via the retained wavelet coefficients — the span of the wavelet basis functions corresponding to the nonzero coefficients is an estimate of $\mathcal{S}$. In short, wavelet shrinkage can be viewed as an approximation to the Wiener filter with $\boldsymbol{U} \approx \boldsymbol{W}$ and $\widetilde{\boldsymbol{\Lambda}} \approx \boldsymbol{H}$ from (5) or (6).

Consider now the wavelet shrinkage procedure of Figure 1, and suppose, without loss of generality, that the first $N_\tau < N$ of the coefficients are retained by the thresholding operation (5) or (6). This can be accomplished by sorting the wavelet coefficients in descending order of magnitude. In the hard shrinkage filter $\boldsymbol{H}_h$, the N_τ weighting coefficients are given by

$$h(i) \;=\; \begin{cases} 1, & i = 1, 2, \ldots, N_\tau \\ 0. & i = N_\tau + 1, N_\tau + 2, \ldots, N, \end{cases} \tag{12}$$

whereas the MMSE-optimal Wiener weights $h_w(i)$ for the N_τ-dimensional subspace determined by the thresholding are given by (7). The sorted weighting coefficients $h_h(i)$ and $h_w(i)$ for the Doppler signal are shown in Figure 2(h).

Assuming a perfect subspace estimate ($N_\tau = N_s$, in particular), the increase in MSE due to the approximation (12) is given by

$$\frac{\text{MSE}_h}{\text{MSE}_w} = \frac{N_\tau}{\sum_{i=1}^{N_\tau} \frac{\theta^2(i)}{\theta^2(i)+\sigma^2}}. \tag{13}$$

Thus, the larger the variation in the signal power (relative to σ^2) over the subspace, the greater the loss in performance due to simple thresholding as opposed to optimal Wiener weighting. In general, smoother signals will have a larger energy spread over the coefficients, resulting in a substantial loss in performance. Experiments with a variety of real and synthetic signals indicate that significant gains in both MSE and visual quality can be achieved with the optimal diagonal weighting.

In the next Section, we present a simple method that uses wavelet shrinkage in a bootstrap fashion to design an empirical wavelet-domain Wiener filter that approaches a weighting profile closer to the optimal one. Such a subspace-based Wiener filter design can yield substantially improved signal estimates as compared to more classical wavelet shrinkage approaches.

3. WAVELET-BASED WIENER FILTERING

Recall our goal of designing a wavelet-domain diagonal Wiener filter $\boldsymbol{H}$ as in Figure 1. Suppose that only N_s (the first N_s, without loss of generality) wavelet coefficients are effectively nonzero, which define the signal subspace $\mathcal{S}$. The diagonal Wiener filter coefficients in the subspace are given by

$$h_w(i) = \frac{\theta^2(i)}{\theta^2(i)+\sigma^2}, \qquad i = 1, 2, \ldots, N_s. \tag{14}$$

To design an empirical Wiener filter, we must estimate $\theta^2(i)$ and σ^2 from the data. Estimating $\theta^2(i)$ poses the challenge, since for sufficiently large N a reliable estimate of σ^2 can be obtained from the finest scale wavelet coefficients.[4,5]

For the $\theta^2(i)$ that are large compared to σ^2, $h_w(i) \approx 1$, and thus, the Wiener filter offers little gain compared to a (hard) thresholded estimate. On the other hand, coefficients $\theta^2(i)$ that are small or comparable in size to σ^2 result in $h_w(i) \ll 1$ and contribute the most to the gain in MSE. This is consistent with the fact the MSE of the Wiener filter

$$\text{MSE}_w = \sigma^2 \sum_{i=1}^{N_s} \frac{\theta^2(i)}{\theta^2(i)+\sigma^2} \tag{15}$$

has the highest sensitivity at $\theta^2(i) = 0$, which decreases monotonically to zero as $\theta^2(i) \longrightarrow \infty$.

Let $N_t < N_s$ be the number of large *trustworthy* signal coefficients $\theta(i)$ for which $h_w(i) \approx 1$, and let $N_d = N_s - N_t$ be the number of small *dubious* signal coefficients. The (hard) thresholding procedure provides an estimate of N_t, since it retains the high SNR (large) wavelet coefficients and zeros most of the noisy coefficients (in the process substantially reducing the variance in the estimate). For the N_t trustworthy coefficients, $h_w(i) \approx 1$, and thus we can use the simple estimate

$$\widehat{\theta}(i) = y(i), \qquad i = 1, 2, \ldots, N_t. \tag{16}$$

However, the N_d dubious signal coefficients suffer from low SNR in $\boldsymbol{y}$, and thus it is difficult to estimate them reliably. To obtain more trustworthy estimates of the dubious coefficients, we take a predictive approach: We *predict* the N_d low SNR coefficients from the N_t high SNR coefficients that can be reliably estimated via (16).

For the prediction process, we propose an indirect approach based on two (slightly) different wavelet transforms, $\boldsymbol{W}_1$ and $\boldsymbol{W}_2$, each of which is appropriate for processing the signal at hand. The procedure is illustrated in Figure 3. The transform $\boldsymbol{W}_1$ is used to obtain a standard wavelet shrinkage signal estimate $\widehat{\boldsymbol{s}}_1$. This estimate is based on the N_t trustworthy wavelet coefficients (assuming hard thresholding). To implicitly obtain a reliable estimate of the entire $N_s = N_t + N_d$ signal coefficients, we transform $\widehat{\boldsymbol{s}}_1$ by $\boldsymbol{W}_2$ to obtain

$$\widehat{\boldsymbol{\theta}}_{21} = \boldsymbol{W}_2 \widehat{\boldsymbol{s}}_1. \tag{17}$$

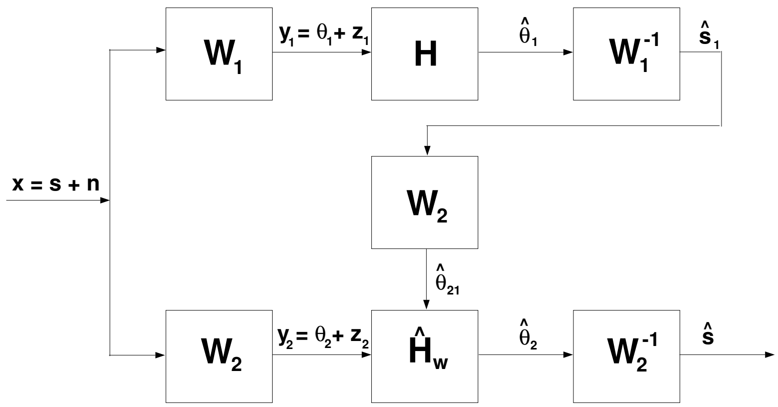

Figure 3. Wavelet-based empirical Wiener filtering. In the upper path, wavelet transform $\boldsymbol{W}_1$ is used to produce the pilot signal estimate $\widehat{\boldsymbol{s}}_1$. This estimate is then used to design an empirical Wiener filter, which is applied to the original noisy signal in the $\boldsymbol{W}_2$ domain.

We call $\widehat{\boldsymbol{s}}_1$ the *pilot estimate*. The estimates $\widehat{\boldsymbol{\theta}}_1$ and $\widehat{\boldsymbol{\theta}}_{21}$ are related via

$$\widehat{\boldsymbol{\theta}}_{21} = \boldsymbol{W}_2 \boldsymbol{W}_1^{-1} \widehat{\boldsymbol{\theta}}_1, \tag{18}$$

with the underlying interpretation that whereas $\widehat{\boldsymbol{\theta}}_1$ contains estimates of the N_t trustworthy coefficients, the composite operator $\boldsymbol{W}_2\boldsymbol{W}_1^{-1}$ smooths $\widehat{\boldsymbol{\theta}}_1$ to predict the remaining N_d dubious coefficients, and thus yields an estimate of the entire $N_s = N_t + N_d$ signal coefficients. The mismatch between $\boldsymbol{W}_1$ and $\boldsymbol{W}_2$ guarantees that the operator $\boldsymbol{W}_2\boldsymbol{W}_1^{-1}$ in (18) spreads or *stretches* the N_t trustworthy coefficients in $\widehat{\boldsymbol{\theta}}_1$ into a larger number of nonzero coefficients in $\widehat{\boldsymbol{\theta}}_{21}$. However, due to the energy compaction property, $\boldsymbol{W}_2\boldsymbol{W}_1^{-1}$ will not overly spread the coefficients, since both $\boldsymbol{W}_1$ and $\boldsymbol{W}_2$ are appropriate for the signal. Figure 2(h) illustrates the result of this implicit smoothing/stretching operation.

The estimate $\widehat{\boldsymbol{\theta}}_{21}$ of the signal coefficients is then used to design an empirical Wiener filter in the $\boldsymbol{W}_2$ domain via

$$\widehat{h}_w(i) \;=\; \frac{\widehat{\theta}_{21}^2(i)}{\widehat{\theta}_{21}^2(i) + \widehat{\sigma}^2}\,. \tag{19}$$

We term the resulting signal estimation procedure *WienerShrink*.

A key question remains: How to choose the transforms $\boldsymbol{W}_1$ and $\boldsymbol{W}_2$? Experience with the technique has shown the final result to be quite insensitive to the choice, so long as both $\boldsymbol{W}_1$ and $\boldsymbol{W}_2$ are appropriate transforms for a classical wavelet shrinkage procedure. We have experimented with transforms employing different filter lengths, shifted versions of the same basis, and even different schemes for dealing with boundary conditions. Current work is aimed at an analytical characterization of the impact of the choice of basis on the performance of the algorithm.

An alternative interpretation of *WienerShrink* is that classical shrinkage overly shrinks the wavelet coefficients (by retaining only the trustworthy coefficients) thereby resulting in a suboptimal weighting, whereas the *WienerShrink* procedure *stretches* the shrunk wavelet coefficients (brings the dubious coefficients back into the picture) to design an improved weighting profile. In this context, our method could alternatively be termed *Asymmetric StretchShrink*.* Several examples of the ASS† algorithm are presented in the next Section.

4. RESULTS

We begin with Donoho's Doppler test signal as illustrated in Figure 2, which shows estimates resulting from hard thresholding, *HybridShrink*, optimal Wiener filtering, and *WienerShrink* (based on both hard and hybrid pilots). In

*In practice, we often use a hard thresholding in the upper branch of Figure 3. Hence, to circumvent any incorrect interpretation, a more appropriate term for the proposed algorithm, by analogy to Donoho and Johnstone's *WaveChop*,[4] would be *WienerChop*.

†Patent pending.

Figure 2(h), we overlay the transfer functions (assuming sorted wavelet data) for the hard threshold, optimal Wiener filter, and two empirical Wiener filters (both designed using the pilot estimate $\widehat{\boldsymbol{s}}_1$ based on hard thresholding in the wavelet basis $\boldsymbol{W}_1$ — see the upper branch of Figure 3). By applying the empirical Wiener filter in the $\boldsymbol{W}_2$ domain (with $\boldsymbol{W}_2 \neq \boldsymbol{W}_1$), we stretch its response to closely resemble the optimal Wiener filter. In contrast, by applying the empirical Wiener filter in the $\boldsymbol{W}_1$ domain (by setting $\boldsymbol{W}_2 = \boldsymbol{W}_1$), the filter cannot escape the signal subspace estimated by the hard thresholding. Zoomed in versions of the estimates, focusing on the initial high-frequency portion, are plotted in Figure 4.

Tables 1–4 record the bias squared, variance, and MSE results for the various estimates on Donoho's four one-dimensional test signals: Doppler, Heavisine, Bumps, and Blocks.[‡] In particular, note how typically *WienerShrink* decreases the bias and variance *simultaneously* as compared to hard thresholding alone. In contrast, the (modified) hard threshold estimate[8] performs better than the *HybridShrink* estimate by *trading* bias for variance. The maximum gain over *HybridShrink* is a factor of 1.95 (for Doppler), while the maximum gain over hard thresholding is a factor of 1.62 (for Bumps).

Table 1. Algorithm performance on Doppler test signal, $\sigma = 1$.

Estimator	MSE	Bias Squared	Variance
Hard threshold	0.209	0.069	0.140
WienerShrink (hard)	0.121	0.016	0.105
HybridShrink	0.274	0.169	0.105
WienerShrink (hybrid)	0.140	0.040	0.100
Ideal Wiener	0.069	0.009	0.060

Table 2. Algorithm performance on Heavisine test signal, $\sigma = 1$.

Estimator	MSE	Bias Squared	Variance
Hard threshold	0.102	0.026	0.076
WienerShrink (hard)	0.075	0.015	0.060
HybridShrink	0.094	0.057	0.037
WienerShrink (hybrid)	0.069	0.033	0.036
Ideal Wiener	0.037	0.008	0.029

Table 3. Algorithm performance on Bumps test signal, $\sigma = 1$.

Estimator	MSE	Bias Squared	Variance
Hard threshold	0.392	0.126	0.266
WienerShrink (hard)	0.242	0.026	0.216
HybridShrink	0.439	0.199	0.240
WienerShrink (hybrid)	0.285	0.023	0.262
Ideal Wiener	0.183	0.026	0.157

Figures 5 and 6 compare the performance of *WienerShrink* to a hard threshold estimate on noisy "Lenna" and "Boy" images. *WienerShrink* improves both the bias, variance, and the visual quality of the images compared to hard thresholding. Similar results hold for soft threshold and *HybridShrink* estimates.

[‡]We used the following bases for $\boldsymbol{W}_1$ in Figure 3. Doppler and Heavisine: Daubechies length-8 most nearly symmetric; Bumps: standard Daubechies length-4; Blocks: Haar. For $\boldsymbol{W}_2$, we employed the Daubechies length-16 most nearly symmetric basis, except for Blocks, where we used the Haar basis. The hard threshold in the upper branch of Figure 3 was set using a modified version of Donoho's rule proposed by Odegard.[8]

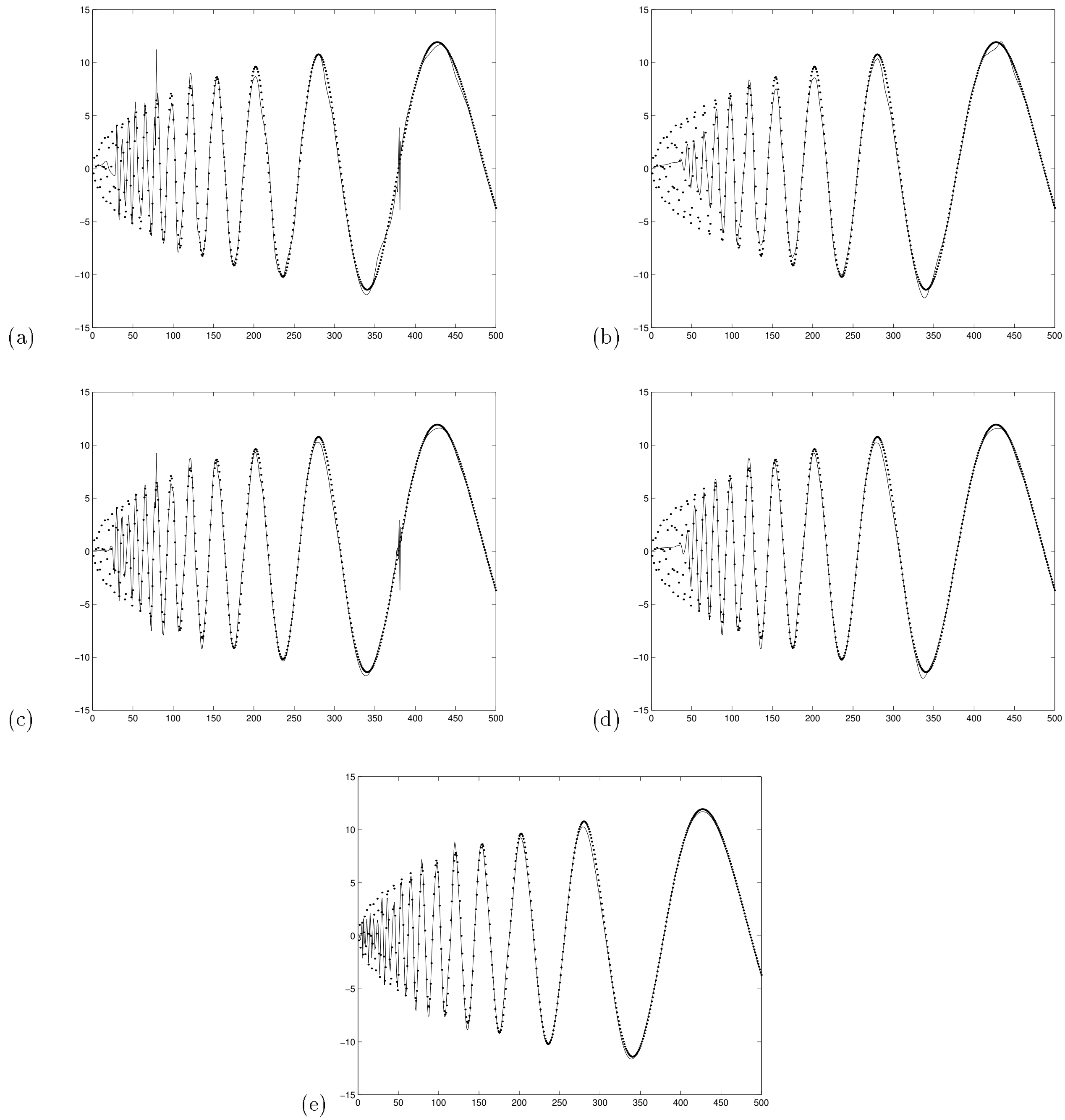

Figure 4. Zooms of the signals from Figure 2. (a) Hard threshold, (b) *HybridShrink*, (c) empirical Wiener filter based on hard threshold pilot estimate., (d) empirical Wiener filter based on *HybridShrink* pilot estimate, (e) optimal Wiener filter. In each plot, the clean signal is shown dotted.

Table 4. Algorithm performance on Blocks test signal, $\sigma = 1$.

Estimator	MSE	Bias Squared	Variance
Hard threshold	0.121	0.011	0.110
WienerShrink (hard)	0.118	0.014	0.104
HybridShrink	0.226	0.085	0.141
WienerShrink (hybrid)	0.167	0.005	0.162
Ideal Wiener	0.064	0.002	0.062

5. CONCLUSIONS

In this paper, we have presented a new scheme for wavelet denoising of signals and images that is a hybrid of standard thresholding and empirical Wiener techniques. The resulting cooperation creates a wavelet-domain filter close to the MSE-optimal Wiener filter. While the *WienerShrink* programme requires the calculation of two wavelet transforms, the increase in performance should outweigh the slight (constant factor) increase in computational cost. We have emphasized designing the empirical Wiener filter using the hard threshold (mainly for analytical and explanative convenience) and *HybridShrink* pilot estimates; however, the *WienerShrink* concept can be based on *any* wavelet shrinkage scheme, including soft thresholding, with similar performance gains.

Current research is directed at an analysis of the choice of $\boldsymbol{W}_1$ and $\boldsymbol{W}_2$ in the hybrid scheme. In particular, we seek to understand how their mismatch affects the performance of the algorithm. The answer may lie in the theory of frames, which have already proved useful for wavelet-based denoising.[9]

Several directions exist for future work on this subject. Nondiagonal Wiener filters can be used to capture the correlations that remain between the wavelet coefficients due to the wavelet transform not being a perfect KL transform. Such filters could be viewed as second-order variations of the work of Crouse et al.[10,11] *WienerShrink* also has potential in the detection, classification, and compression arenas. In particular, the subspace-based interpretation of *WienerShrink* can be exploited in the *estimator-correlator*[6] formulation of optimal detectors.

Acknowledgements: The authors thank Matthew Crouse for discussions on wavelet denoising and Jazzy Jones and Akira for pointing out the significance of the JAM in this context.

REFERENCES

1. D. L. Donoho, "Unconditional bases are optimal bases for data compression and for statistical estimation," *App. Comp. Harmonic Anal.* **1**, pp. 100–115, Dec 1993.
2. I. Daubechies, *Ten Lectures on Wavelets*, SIAM, New York, 1992.
3. M. Vetterli and J. Kovačević, *Wavelets and Subband Coding*, Prentice-Hall, Englewood Cliffs, NJ, 1995.
4. D. L. Donoho and I. Johnstone, "Adapting to unknown smoothness via wavelet shrinkage," *J. Amer. Stat. Assoc.* **90**, pp. 1200–1224, Dec. 1995.
5. D. Donoho, "De-noising by soft-thresholding," *IEEE Trans. Inform. Theory* **41**, pp. 613–627, May 1995.
6. H. V. Poor, *An Introduction to Signal Detection and Estimation*, Springer-Verlag, 1988.
7. R. D. Nowak, "Optimal signal estimation using cross-validation," *IEEE Signal Processing Letters* **4**, pp. 23–25, Jan. 1997.
8. J. E. Odegard. Personal communication.
9. M. Lang, H. Guo, J. E. Odegard, C. S. Burrus, and R. O. Wells, "Nonlinear processing of a shift invariant DWT for noise reduction," 1995. *SPIE Symp. OE/Aerospace Sensing and Dual Use Photonics*, Orlando, FL, 17-21 April 1995.
10. M. S. Crouse, R. D. Nowak, and R. G. Baraniuk, "Hidden markov models for wavelet-based signal processing," in *Proc. 30th Asilomar Conf.*, (Pacific Grove, CA), 1997.
11. M. S. Crouse, R. D. Nowak, and R. G. Baraniuk, "Wavelet-based statistical signal processing using hidden markov models," *IEEE Trans. Signal Processing* , 1997. Submitted.

(a) (b)

(c) (d)

Figure 5. Image denoising using wavelet shrinkage I. (a) 512×512 Lenna image. (b) Lenna corrupted by white Gaussian noise with $\sigma = 20$. (c) Hard threshold estimate (ensemble-averaged statistics: MSE=92.44, bias squared=49.17, variance=43.27) (d) *WienerShrink* estimate (MSE=61.37, bias squared=33.99, variance=27.38). Gain in MSE=1.51.

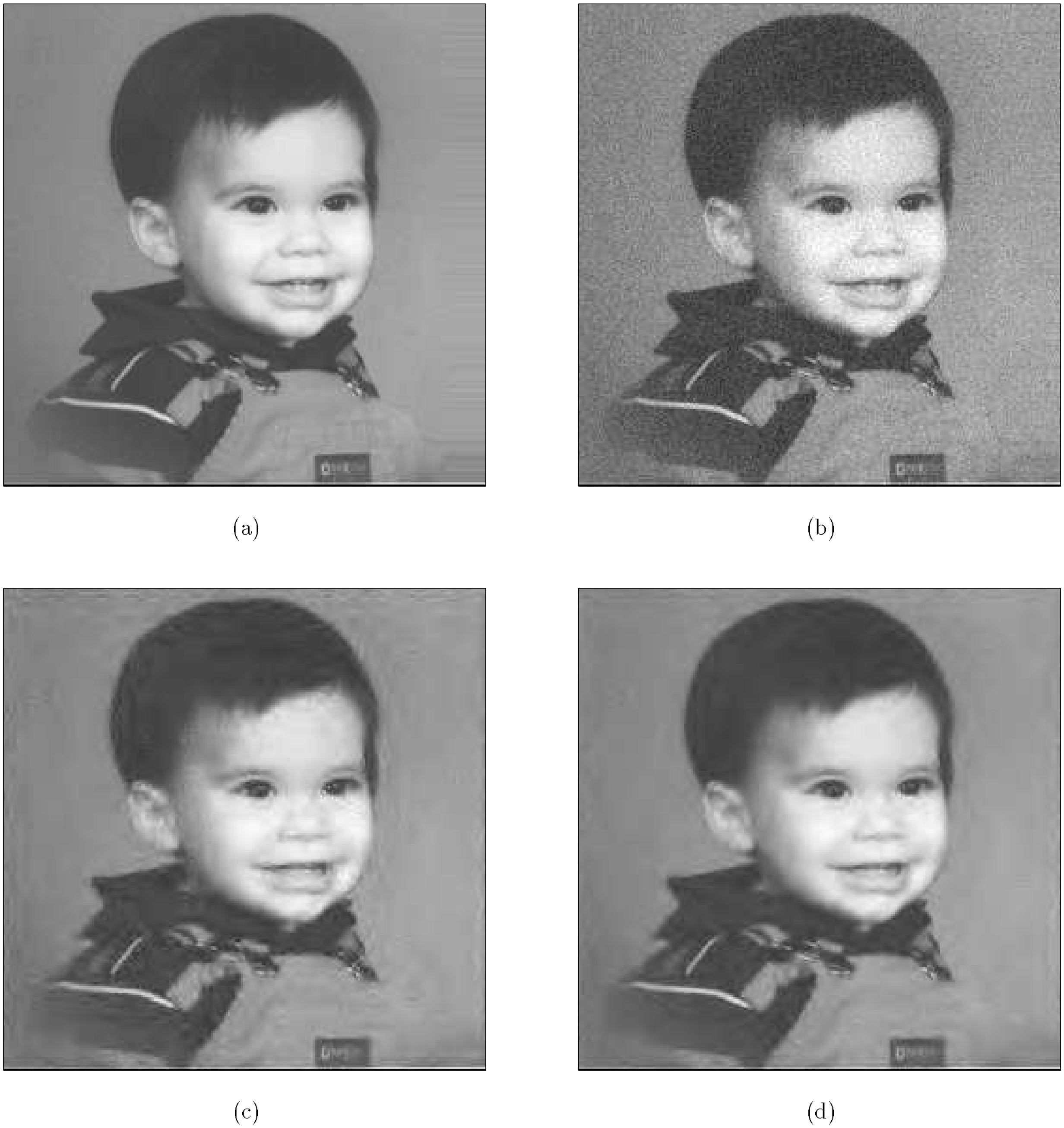

Figure 6. Image denoising using wavelet shrinkage II. (a) 256×256 Boy image. (b) Boy corrupted by white Gaussian noise with $\sigma = 10$. (c) Hard threshold estimate (ensemble-averaged statistics: MSE=74.93, bias squared=56.04, variance=18.89). (d) *WienerShrink* estimate (MSE=48.07, bias squared=36.66, variance=11.41). Gain in MSE=1.56.