ANALYSIS OF WAVELET–DOMAIN WIENER FILTERS

Hyeokho Choi and Richard Baraniuk

Department of Electrical and Computer Engineering, Rice University, Houston, TX 77251–1892, USA

ABSTRACT

We investigate Wiener filtering of wavelet coefficients for signal denoising. Empirically designed wavelet-domain Wiener filters outperform many other denoising algorithms based on wavelet thresholding. However, up to now, it has not been clear how to choose the signal model used to design the filter, because the effect of model selection on the filter performance is difficult to understand. By analyzing the error involved in the Wiener filter designed with an empirically obtained signal model, we show that hard thresholding is typically outperformed by a Wiener filter designed in an alternate wavelet domain. Our analysis furthermore provides a method for selecting the various parameters involved in a wavelet-domain Wiener filtering scheme.

1. INTRODUCTION

Denoising algorithms attempt to recover a signal corrupted by additive white noise. The signal we consider can be modeled as a vector in $\mathbf{R}^N$ space.[1] The noisy signal can be written as $s = x + n$, where x and n model signal and noise, respectively. Let $s(i)$ denote the i-th sample of s, and define $x(i)$ and $n(i)$ similarly. Let $\mathbf{W}$ denote a wavelet transform [1] that is well matched to the signal under consideration. By transforming the observed noisy signal into the $\mathbf{W}$ wavelet domain, we obtain wavelet coefficients $y = \theta + z$, where $y = \mathbf{W}s$, $\theta = \mathbf{W}x$ and $z = \mathbf{W}n$, respectively. Thanks to the compaction and decorrelation properties of wavelet transforms [1], we can devise many filtering algorithms that estimate the true signal from the noisy observation. Many algorithms that threshold or shrink the wavelet coefficients and then inverse transform to the time domain have been proposed [2, 3].

The idea of Wiener filtering of individual wavelet coefficient arises from the fact that wavelet transforms tend to decorrelate real-world signals. That is, the wavelet transform approximates the Karhunen-Loève transform. To recover θ from y, Wiener filtering of individual wavelet coefficients is optimal in the sense of minimizing the mean-square error (MSE) (assuming perfect decorrelation of noisy wavelet coefficients). The Wiener filtering of each wavelet coefficient is given as $\tilde{\theta}(i) = \frac{\theta^2(i)}{\theta^2(i)+\sigma^2}y(i)$, where σ^2 is the variance of $z(i)$ [1, 4]. Because the $\theta(i)$'s are unknown, we use estimated values $\hat{\theta}(i)$ instead, and we obtain the empirical Wiener

[1] While our ideas apply without modification to higher-dimensional data such as images, for notational convenience we will focus on 1-d signals.

filter

$$\tilde{\theta}(i) = \frac{\hat{\theta}^2(i)}{\hat{\theta}^2(i) + \sigma^2} y(i). \tag{1}$$

Figure 1 illustrates the improved wavelet-domain Wiener denoising technique (*WienerChop*) of [5]. The Wiener filtering is applied in the $\mathbf{W_2}$ domain, and the signal model ($\hat{\theta}$) needed to design the filter is obtained by hard-threshold denoising the signal in the $\mathbf{W_1}$ domain. It was shown in [5] that this empirical Wiener filtering algorithm outperforms many denoising algorithms using thresholding or shrinkage. In this paper, we will analyze this and related denoising algorithms in terms of optimal filtering of noisy wavelet coefficients.

Many other wavelet thresholding and shrinkage schemes can be shown to be approximate forms of Wiener filtering. Thus, it is worthwhile to consider these wavelet denoising algorithms from the viewpoint of Wiener filtering. In this way, we can analyze many wavelet denoising algorithms with a unified approach.

We aim to analyze the errors involved in the Wiener filtering of wavelet coefficients using a filter designed based on specific signal models obtained through other methods. In particular, our main contribution is the analysis of the denoising mechanism of the *WienerChop* of [5]. Although the *WienerChop* algorithm is superior to many other denoising algorithms, the behavior of the algorithm is not clearly understood. Our analysis will clearly indicate the main source of improvement in the performance of the empirical Wiener filtering over other denoising algorithms using thresholding. We will show that the superiority of the *WienerChop* algorithm is mainly due to the reduction of the error that results from the mismatch between the empirical signal model and the true signal. The success of the algorithm comes from the proper reconditioning of the signal model by an orthonormal transformation to a different domain. Our analysis will furthermore suggest methods to design good wavelet-domain Wiener filters.

2. ERRORS IN EMPIRICAL WIENER FILTERING

The signal estimation error when we use an approximate Wiener filter in (1) warrants further consideration. We can obtain the expression for the MSE as $\text{MSE} = E_{\text{opt}}(\theta) + E_{\text{mis}}(\theta, \hat{\theta})$, where

$$E_{\text{opt}}(\theta) = \sum_{i=1}^{N} \frac{\theta^2(i)\sigma^2}{\theta^2(i) + \sigma^2}, \tag{2}$$

$$E_{\text{mis}}(\theta, \hat{\theta}) = \sum_{i=1}^{N} (\theta^2(i) + \sigma^2) \left(\frac{\hat{\theta}^2(i)}{\hat{\theta}^2(i) + \sigma^2} - \frac{\theta^2(i)}{\theta^2(i) + \sigma^2} \right)^2. \tag{3}$$

The term E_{opt} represents the MSE of the Wiener filter when we have $\hat{\theta}(i) = \theta(i)$ for all i; E_{mis} is the error resulting from the mismatch of the signal model to the true signal.

Once we fix the wavelet transform, E_{opt} is determined for a given signal. E_{opt} attains its minimum value when the signal energy is concentrated on a single wavelet coefficient $\theta(j)$ for some j and $\theta(i) = 0$ for $i \neq j$. In other words, the optimal Wiener filter performs well when the wavelet transform compacts the signal well.

The term $E_{\mathrm{mis}}(\theta, \hat{\theta})$ represents the error resulting from mismatch of the signal model $\hat{\theta}$ with the signal θ. When the signal model is perfect, $E_{\mathrm{mis}} = 0$, and we obtain the performance of the optimal Wiener filter. However, in practice we do not know the exact values of $\theta(i)$'s, and this error term is not zero. To see how this error term varies as $\hat{\theta}(i)$ changes, consider the behavior of each term in the expression for $E_{\mathrm{mis}}(\theta, \hat{\theta})$ in (3). Figure 2 shows the plot of a term in the summation for $\sigma = 0.1$, $\theta(i) \in [-0.5, 0.5]$, and $\hat{\theta}(i) \in [-0.5, 0.5]$. E_{mis} is the summation of N such terms. In this figure, we see that a relatively large error can result if $\hat{\theta}(i) = 0$ when $\theta(i)$ is not small. For a fixed $\theta(i)$, the error is minimized when $\hat{\theta} = \theta$. When θ is close to zero, misestimation of θ has little effect on the MSE. Also, if $\hat{\theta}$ increases away from zero, then the MSE decreases rapidly even when θ is not small. In these two cases, the value approaches σ^2.

3. INTERPRETATION OF HARD THRESHOLDING

We can interpret many wavelet-domain denoising algorithms as special cases of Wiener filters designed with some signal model in the wavelet domain. As a preparation for the arguments in the following sections, we consider the hard thresholding scheme [3] in view of the error analysis of Section 2. With the model of the noisy wavelet coefficients after transforming the signal by $\mathbf{W_1}$ given by $y(i) = \theta(i) + z(i)$, $i = 1, \cdots, N$, hard thresholding of wavelet coefficients can be viewed as an approximate form of Wiener filtering:

$$\hat{y}(i) = \frac{\hat{\theta}^2(i)}{\hat{\theta}^2(i) + \sigma^2} y(i), \quad \text{with } \hat{\theta}(i) = \begin{cases} \infty & \text{if } |y(i)| > \tau \\ 0 & \text{otherwise,} \end{cases}$$

(4)

where τ is a threshold determined according to the noise variance. In other words, $\hat{\theta}(i)$ in (4) is the signal model used to design the filter, and under this model, the Wiener filter is the hard thresholding algorithm.

To see how much error would result from hard thresholding, we use the formulas for MSE derived in Section 2. The term E_{opt} is fixed once the underlying wavelet transform is determined. In view of Fig. 2, we see that E_{mis} mainly consists of the error resulting from $\hat{\theta}(i) \neq \theta(i)$ for those wavelet coefficients $|\theta(i)| \simeq \tau$ and $|y(i)| < \tau$. Such coefficients are removed by hard thresholding, and we have $\hat{\theta}(i) = 0$, making the corresponding error term in E_{mis} (see (3)) large. Hard thresholding has a poor performance in this sense.

Other wavelet thresholding or shrinkage algorithms can be interpreted as Wiener filters in a similar way. In particular, we expect similar behavior of E_{mis} for many algorithms, including soft thresholding, where small wavelet coefficients are also removed.

In [1, pp. 425-464], a similar comparison between hard thresholding and the optimal Wiener filtering (ideal attenuation of wavelet coefficients) was made by analyzing the errors involved in each algorithm.

4. ANALYSIS OF THE WIENERCHOP ALGORITHM

Although wavelet-domain denoising using hard thresholding can be used to estimate a signal, this estimated signal can also be used as a signal model to *design a filter* as in (1). Then, we can process the original noisy signal using this filter to obtain a better estimate of the signal, because the signal model used to design the filter (obtained by hard thresholding the wavelet coefficients in $\mathbf{W_1}$) may be better matched to the signal than the model in (4). However, using the signal model obtained by hard thresholding in $\mathbf{W_1}$ to design a Wiener filter in the same domain experiences similar problems (large E_{mis}) as simple hard thresholding. Rather, we can think of Wiener filtering in an alternate domain by an orthonormal transformation.

The *WienerChop* algorithm of [5] follows from this idea. The signal model is again provided by wavelet-domain hard thresholding using the wavelet transform $\mathbf{W_1}$. Although an orthonormal transformation does not change the mean-square error of signal model $E(|\hat{\theta} - \theta|^2)$ in the new domain, the signal model in the transformed domain may be more suitable for designing a Wiener filter, reducing the error E_{mis} due to model mismatch in the design. In particular, we can avoid the type of model mismatch that can cause large errors (as in hard thresholding) by a coordinate transformation, and this may significantly reduce the overall error.

Let $\mathbf{W_1}$ be the wavelet transform used to obtain the signal model using hard thresholding, and let $\mathbf{K}$ be an orthonormal transformation from the $\mathbf{W_1}$ domain to a different one. Let $\mathbf{H}$ be the hard thresholding operator defined as $\mathbf{H} = \mathrm{diag}[h(1), h(2), \cdots, h(N)]$, where $h(i) = 1$ if $|y(i)| > \tau$ and $h(i) = 0$ otherwise, and τ is the threshold. Then, the signal model obtained by hard thresholding in the $\mathbf{W_1}$ domain is $\mathbf{y_h} = \mathbf{Hy} = \mathbf{HW_1 s}$. The estimated signal in the time domain can be written as $\mathbf{s_h} = \mathbf{W_1^{-1} y_h} = \mathbf{W_1^{-1} HW_1 s}$.

Suppose $\mathbf{K}$ is an orthonormal transformation from $\mathbf{W_1}$ coordinates to a new coordinates. Depending on the choice of this transformation, we can expect that the unfavorable mismatches of the signal model may be mitigated, rendering a more favorable signal model so that the Wiener filter designed in the new domain can have significant reduction of error. The noisy signal and the signal model in the new domain are given by $\mathbf{y_k} = \mathbf{Ky} = \mathbf{KW_1 s}$ and $\mathbf{y_{hk}} = \mathbf{Ky_h} = \mathbf{KHW_1 s}$. We can design a Wiener filter in the new domain based on the signal model $\mathbf{y_{hk}}$, and then the filter can be applied to $\mathbf{y_k}$. We obtain the final time-domain signal by inverse transforming back to the time domain.

When choosing $\mathbf{K}$, we should be careful that the resulting transform of the signal in the new domain ($\mathbf{y_k}$) is as compact as possible, so that the inherent Wiener filtering error (E_{opt}) is small. In this respect, it is desirable to choose $\mathbf{K}$ so that the resulting new domain is another wavelet domain. Let $\mathbf{W_2}$ denote this wavelet domain. Then, $\mathbf{K}$ has the form $\mathbf{K} = \mathbf{W_2 W_1^{-1}}$. In terms of $\mathbf{W_2}$, the signal is represented as $\mathbf{y_2} = \mathbf{W_2 s}$ in the $\mathbf{W_2}$ domain, and we have the signal model $\mathbf{y_{h2}} = \mathbf{W_2 W_1^{-1} HW_1 s}$. This corresponds to having $\mathbf{W_1} \neq \mathbf{W_2}$ in the algorithm of [5]. When $\mathbf{W_1} \neq \mathbf{W_2}$,

614

the error E_{mis} incurred by Wiener filtering in the $\mathbf{W_2}$ domain depends on the ability of the transform $\mathbf{K} = \mathbf{W_2}\mathbf{W_1}^{-1}$ to spread the model mismatches in the $\mathbf{W_1}$ domain by transforming into the $\mathbf{W_2}$ domain. This makes the model more suitable for designing a Wiener filter in the $\mathbf{W_2}$ domain, assuming that the signal is compactly represented in both $\mathbf{W_2}$ and $\mathbf{W_1}$.

In view of the idea of mitigation of signal modeling error to reduce E_{mis}, we can pick a good pair of wavelet transforms $\mathbf{W_1}$ and $\mathbf{W_2}$ for a given signal. However, because the original signal is unknown, it is hard to characterize the influence of these wavelet bases on the estimation error.

When we have more than two wavelet bases under which the signal has compact representations, we can consider an iterative scheme of choosing a pair of wavelets at a time. The signal estimated from the Wiener filter can again be used as a signal estimate to design yet another Wiener filter in different wavelet domain. However, because the estimation error of the empirical Wiener filtering varies very nonlinearly as the signal model changes, we are not guaranteed to improve performance. Our analysis has shown that we can obtain improvement in estimation error compared with hard thresholding when we use the hard thresholded signal estimate to design the Wiener filter. Thus, iterating the empirical Wiener filtering for multiple wavelet bases is not guaranteed to converge to a good signal estimate.

5. DENOISING EXAMPLE

This idea of mitigating the effect of unfavorable model errors by an orthonormal transformation is well illustrated by an example. Figure 3 shows the test signal (obtained by concatenating Donoho's Doppler and Blocks signals [3]) and the same signal corrupted by white Gaussian noise with variance $\sigma = 0.1$. Figure 4 shows the estimated signals using wavelet-domain denoising obtained by hard thresholding in the Haar (D_2) wavelet domain and by empirical Wiener filtering (using $\mathbf{W_1}$ = the Haar wavelet transform and $\mathbf{W_2}$ = the Daubechies length-12 (D_{12}) wavelet transform). In terms of $\mathbf{W_1}$ and $\mathbf{W_2}$, we expect that the compaction of the wavelet coefficients of our test signal will be almost same, because each of these wavelet bases compacts half of the signal very well.

Table 1 shows the error term E_{opt} for different choices of wavelet bases. We note that the values of E_{opt} are similar for both the D_2 and D_{12} bases, because the compaction of the wavelet coefficients is almost the same in each domain. To see how the estimates of the wavelet coefficients and the actual coefficients distribute, we computed $\hat{\theta} = \mathbf{H}\mathbf{W_1}\mathbf{y}$ and $\theta = \mathbf{W_1}\mathbf{x}$. Figure 5 show the distribution of $(\theta(i), \hat{\theta}(i))$ for $i = 1, \cdots, N$. We see that the coefficients with small magnitudes concentrate around the line $\hat{\theta} = 0$, which results in large E_{mis} according to the plot Fig. 2. Other coefficients are gathered along the line $\hat{\theta} = \theta$.

To see the distribution of the coefficients and estimates after transforming into the $\mathbf{W_2}$ domain, we computed $\mathbf{W_2}\mathbf{W_1}^{-1}\hat{\theta}$ and $\mathbf{W_2}\mathbf{x}$. In Fig. 6 we see that there are many fewer points falling around the line $\hat{\theta} = 0$, resulting in much lower E_{mis}. This is because the model mismatches in $\mathbf{W_1}$ domain, which are distributed very unfavorably, were transformed to $\mathbf{W_2}$ domain where they distribute much more favorably. By actually computing E_{mis} with the model obtained in the $\mathbf{W_1}$ domain by hard thresholding and the signal in the same domain, we obtained the values shown in

Table 1 for different choices of wavelet bases. We observe a significant reduction in E_{mis} when we choose $\mathbf{W_1} \neq \mathbf{W_2}$. This reduction in E_{mis} explains the superiority of the algorithm in [5] over the simple hard thresholding algorithm.

6. CONCLUSIONS

In this paper, we have analyzed the errors involved in wavelet-domain empirical Wiener filtering, including the *WienerChop* algorithm of [5]. We showed that errors due to the mismatch of signal model can be significant. Fortunately, this error can be reduced by transforming both the signal and model to another wavelet domain where the Wiener filter will yield a smaller error.

The difficulty in the analysis of the wavelet-domain Wiener filters arises from the signal dependence of the processing. The influence of the choice of $\mathbf{W_1}$ and $\mathbf{W_2}$ on the overall performance is hard to analyze in general. The development of an algorithm to design the wavelet transforms for a given signal remains a task for future research.

We have recently learned about another method for denoising using multiple wavelet domains [6]. Currently, we are investigating filtering algorithms to incorporate more than two wavelet transforms to obtain improved performance.

7. REFERENCES

[1] S. Mallat, *A Wavelet Tour of Signal Processing*, Academic Press, 1998.

[2] D. L. Donoho and I. Johnstone, "Adapting to unknown smoothness via wavelet shrinkage," *J. Amer. Stat. Assoc.* pp.1200-1224, Dec. 1995.

[3] D. L. Donoho, "De-noising by soft-thresholding," *IEEE Trans. Inform. Theory*, vol. 41, no. 3, pp. 613-627, May 1995.

[4] H. V. Poor, *An Introduction to Signal Detection and Estimation*, Springer Verlag, 1988.

[5] S. P. Ghael, A. M. Sayeed and R. G. Baraniuk, "Improved Wavelet Denoising via Empirical Wiener Filtering," *Proceedings of SPIE*, vol. 3169, pp. 389-399, San Diego, July 1997.

[6] P. Ishwar, K. Ratakonda, P. Moulin, and N. Ahuja, "Image denoising using multiple compaction domains," *Proceedings of ICASSP '98*, pp. 1889-1892, Seattle, 1998.

Table 1: Error terms depending on the choice of wavelet bases $\mathbf{W_1}$ and $\mathbf{W_2}$. Choosing $\mathbf{W_1} \neq \mathbf{W_2}$ reduces E_{mis} significantly.

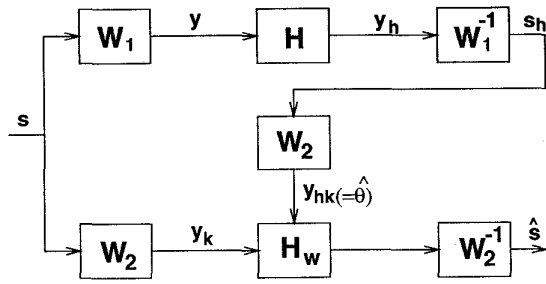| $\mathbf{W_1}$	D_2	D_{12}	D_2	D_{12}
$\mathbf{W_2}$	D_2	D_{12}	D_{12}	D_2
E_{opt}	2.2165	2.0610	2.0610	2.2165
E_{mis}	3.4443	2.6966	0.7748	0.5843
MSE	5.6608	4.7576	2.8358	2.8008

Figure 1: Wavelet-based empirical Wiener filtering. In the upper path, wavelet transform $\mathbf{W_1}$ is used to produce a signal model $\mathbf{y_{hk}}(=\hat{\theta})$. This model is then used to design an empirical Wiener filter $\mathbf{H_w}$ that is applied to the original noisy signal in the $\mathbf{W_2}$ domain.
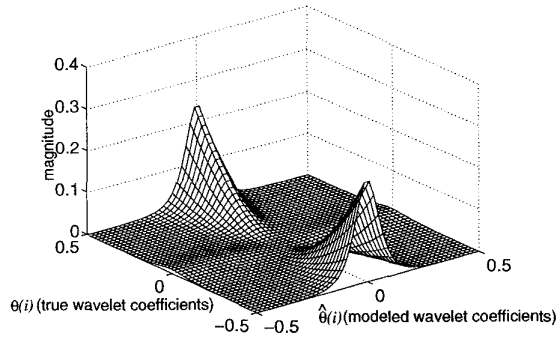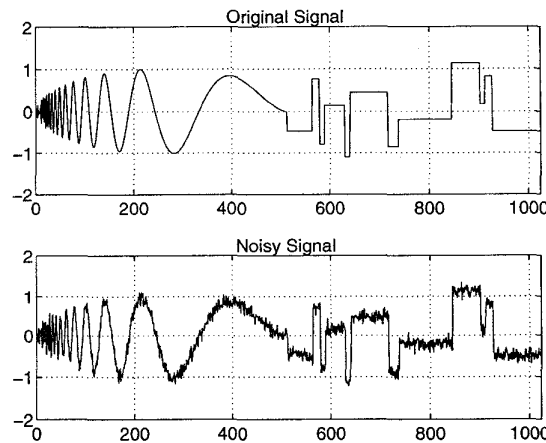

Figure 2: Error due to model mismatch.

Figure 3: Above: DoppleBlock test signal. Below: signal corrupted by additive white Gaussian noise with $\sigma = 0.1$.
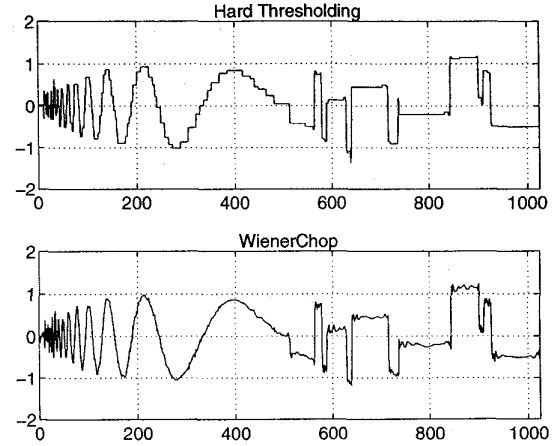

Figure 4: Denoised signals. Above: hard thresholding in D_2 domain, MSE=5.6608. Below: *WienerChop* using $\mathbf{W_1} = D_2$ and $\mathbf{W_2} = D_{12}$, MSE=2.8358.
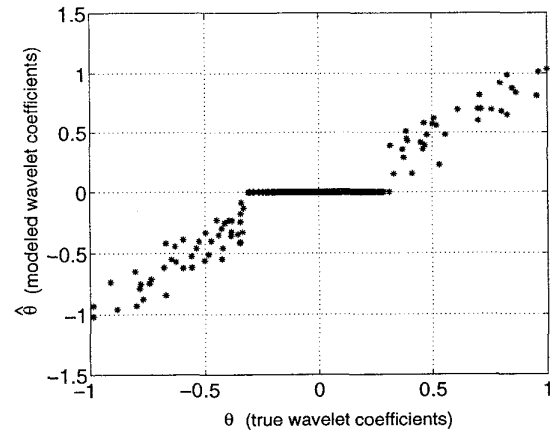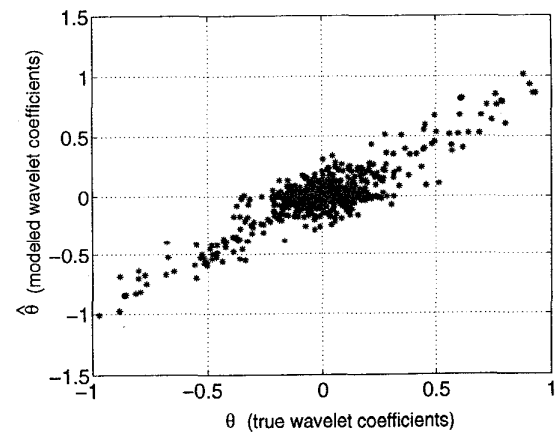

Figure 5: Model mismatch in the $\mathbf{W_1}$ domain.

Figure 6: Model mismatch in the $\mathbf{W_2}$ domain.

616