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 have superior performance over other denoising algorithms using wavelet thresholding. However, it is not clear how we should choose the signal model that is used to design the filter, because the effect of model selection on the filter performance was hard to understand. We analyze the error involved in the Wiener filters designed with an empirically obtained signal model showing that we can almost always obtain better performance than hard thresholding algorithm by using Wiener filters in another wavelet domain with the signal model obtained by hard thresholding. Our analysis furthermore provides a method to choose various parameters in 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 $\mathbf{R}^{\mathbf{N}}$ space. The noisy signal can be written as $\mathbf{s} = \mathbf{x} + \mathbf{n}$, where $\mathbf{x}$ and $\mathbf{n}$ model signal and noise, respectively. Let $s(i)$ denote i-th sample of $\mathbf{s}$, and define $x(i)$ and $n(i)$ similarly. Let $\mathbf{W}$ denote a wavelet transform that is well matched to the signal in consideration. By transforming the observed noisy signal into $\mathbf{W}$ wavelet domain, we obtain wavelet coefficients $\mathbf{y} = \theta + \mathbf{z}$, where $\mathbf{y} = \mathbf{W}\mathbf{s}$, $\theta = \mathbf{W}\mathbf{x}$ and $\mathbf{z} = \mathbf{W}\mathbf{n}$, respectively. Thanks to the compaction and decorrelation properties of wavelet transformations [1], we can devise many filtering algorithms that estimate the true signal from a noisy observation. Many algorithms that obtain the estimates of the signal by thresholding or shrinking of wavelet coefficients and then inverse transforming to 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 data. That is, the wavelet transform approximates the Karhunen-Loeve (KL) transform. To recover θ from $\mathbf{y}$, Wiener filtering of individual wavelet coefficient is optimal in the sense of minimizing 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 $\theta(i)$'s are unknown, we use estimated values $\hat{\theta}(i)$ instead, and we obtain an empirical Wiener filter given as

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

This work was supported by the National Science Foundation, grant no. MIP–9457438, DARPA, grant no. DARPA/AFOSR F49620–97–1–0513, and Texas Instruments.

Email: choi@ece.rice.edu, richb@rice.edu

Web: http://www.dsp.rice.edu

Recently, an improved wavelet domain denoising technique has been proposed that utilizes the Wiener filtering of wavelet coefficients [5]. The algorithm is illustrated in Fig. 1. The Wiener filtering is carried out in the $\mathbf{W_2}$ domain, and the signal model ($\hat{\theta}$) needed to design the filter was obtained by denoising the signal in $\mathbf{W_1}$ domain by hard thresholding and then transforming to $\mathbf{W_2}$ domain. It was shown that this empirical Wiener filtering algorithm outperforms many denoising algorithms using thresholding or shrinkage of wavelet coefficients. This idea of wavelet-domain Wiener filtering motivates the analysis of many 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.

Our goal is to analyze the errors involved in the Wiener filtering of wavelet coefficients using a filter designed with specific signal models obtained through some other methods. In particular, our main contribution is the analysis of the denoising mechanism of the empirical Wiener filtering algorithm (*WienerChop*) given in [5]. Although the *WienerChop* algorithm is superior to many other denoising algorithms, the behavior of the algorithm was not clearly understood in [5]. We clearly illustrate the main source of improvement in performance of the empirical Wiener filtering over other denoising algorithms using thresholding. We show that the superiority of the *WienerChop* algorithm is mainly due to the reduction of the error that results from the mismatch of the empirical signal model to 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 can suggest methods to design good wavelet domain Wiener filtering schemes.

2. ERRORS IN EMPIRICAL WIENER FILTERING

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

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

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

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, $E_{\rm opt}$ is determined for a given signal. We can see that $E_{\rm opt}$ attains its minimum when the signal energy is concentrated at a single wavelet coefficient $\theta(j)$ for some j and $\theta(i) = 0$ for $i \neq j$. In other words, the we obtain a good performance of the optimal Wiener filter when the wavelet transform compacts the signal well.

The term $E_{\rm mis}(\theta, \hat{\theta})$ represents the error resulting from mismatch of the signal model $\hat{\theta}$ with the signal θ. When the signal model is perfect, $E_{\rm mis} = 0$ and we obtain the performance of the optimal Wiener filter. However, 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 $E_{\rm mis}(\theta, \hat{\theta})$ 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 $\hat{\theta}(i) \in [-0.5, 0.5]$. $E_{\rm mis}$ is 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 $\theta(i)$, it is minimized when $\hat{\theta} = \theta$. When θ is close to zero, misestimation of θ has little effect on MSE. Also, if $\hat{\theta}$ increases away from zero, 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 filtering scheme designed with some signal model in the wavelet domain. As a preparation for the argument 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 as $y(i) = \theta(i) + z(i)$, $i = 1, \cdots, N$, hard thresholding of wavelet coefficients can be viewed as an approximate form of Wiener filtering as

$$\hat{y}(i) = \frac{\hat{\theta}^2(i)}{\hat{\theta}^2(i) + \sigma^2} y(i), \quad \text{with} \quad \hat{\theta}(i) = \begin{cases} \infty & \text{if } |y(i)| > \tau \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

In other words, $\hat{\theta}(i)$ in (4) is the model of the signal used to design the filter, and under this model, the designed 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 $E_{\rm opt}$ is fixed once the underlying wavelet transform is determined. In view of Fig. 2, we see that $E_{\rm mis}$ 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 $E_{\rm mis}$ (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. And, we expect similar behavior of $E_{\rm mis}$ in many other algorithms including the soft thresholding scheme 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 filter the original 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 $E_{\rm mis}$) as simple hard thresholding does. Rather, we can think of Wiener filtering in an alternate domain by an orthonormal transformation.

The *WienerChop* algorithm in [5] follows from this idea. The signal model is again provided by wavelet domain hard thresholding using 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 due to model mismatch ($E_{\rm mis}$) in the design. In particular, we can avoid the type of model mismatch that can cause possibly large error (as in hard thresholding), by a coordinate transformation, and this may significantly reduce the 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 $\mathbf{W_1}$ 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. Then, the signal model obtained by hard thresholding in $\mathbf{W_1}$ domain is given as $\mathbf{y_h} = \mathbf{Hy} = \mathbf{HW_1s}$. The estimated signal in time domain can be written as $\mathbf{s_h} = \mathbf{W_1^{-1}y_h} = \mathbf{W_1^{-1}HW_1s}$.

Suppose $\mathbf{K}$ is an orthonormal transformation from $\mathbf{W_1}$ coordinates to a new coordinates. Depending on the choice of this orthonormal transformation, we can expect that the unfavorable mismatches of the signal model may be mitigated after coordinate transformation, rendering a more favorable signal model so that the Wiener filter designed in the new domain can have significant reduction of error. Then, the noisy signal and the signal model in the new domain are given as $\mathbf{y_k} = \mathbf{Ky} = \mathbf{KW_1s}$, and $\mathbf{y_{hk}} = \mathbf{Ky_h} = \mathbf{KHW_1s}$. 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 ($E_{\rm opt}$) 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 $\mathbf{K} = \mathbf{W_2W_1^{-1}}$. In terms of $\mathbf{W_2}$, the signal is represented as $\mathbf{y_2} = \mathbf{W_2s}$ in the $\mathbf{W_2}$ domain, and we have the signal model given by $\mathbf{y_{h2}} = \mathbf{W_2W_1^{-1}HW_1s}$. This corresponds to having $\mathbf{W_1} \neq \mathbf{W_2}$ in the algorithm of [5]. When $\mathbf{W_1} \neq \mathbf{W_2}$, the error $E_{\rm mis}$ of the Wiener filtering in $\mathbf{W_2}$ domain depends on the ability of the transform $\mathbf{K} = \mathbf{W_2W_1^{-1}}$ to spread the model mismatches in $\mathbf{W_1}$ domain by transforming into $\mathbf{W_2}$ domain, to make the model suitable for designing a Wiener filter in $\mathbf{W_2}$ domain, assuming the signal is compactly represented in $\mathbf{W_2}$ as well as in $\mathbf{W_1}$.

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 by choosing a pair of wavelets at a time. The signal estimate out of 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 have an improvement in 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 original noiseless signal obtained by concatenating Donoho's Doppler and Blocks signals [3] and the 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 D_2 domain and empirical Wiener filtering. For the signal by the empirical Wiener filtering in Fig. 4, we chose $\mathbf{W_1}$ to be Haar wavelet (D_2) and $\mathbf{W_2}$ to be Daubechies length-12 wavelet (D_{12}), respectively, as wavelet bases. In terms of $\mathbf{W_1}$ and $\mathbf{W_2}$, we expect the compaction of wavelet coefficients of our test signal will be almost same because each of these wavelet bases compacts half of the signal very well while the compaction of the other half is not good. 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 D_2 and D_{12} bases, because the compaction of wavelet coefficients is almost the same in either domains. To see how the estimates of wavelet coefficients and the actual coefficients distribute, we computed $\hat{\theta} = \mathbf{HW_1y}$ and $\theta = \mathbf{W_1x}$. Figure 5 show the distribution of $(\theta(i), \hat{\theta}(i))$ for $i = 1, \cdots, N$. We see that the coefficients with small magnitudes distribute 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 $\mathbf{W_2}$ domain, we computed $\mathbf{W_2W_1^{-1}}\hat{\theta}$ and $\mathbf{W_2x}$ and plotted the result in Fig. 6. In this figure we see that there are much less points falling around the line $\hat{\theta} = 0$, resulting in much less E_{mis}. This is because the model mismatches in $\mathbf{W_1}$ domain, which distributed very unfavorably, was transformed to $\mathbf{W_2}$ domain and they distribute much more favorably in the $\mathbf{W_2}$ domain. 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 obtain 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 simple hard thresholding algorithm.

6. CONCLUSIONS

In this paper, we analyzed the errors involved in wavelet domain empirical Wiener filter. We showed that the error due to the mismatch of signal model can be significant depending on the methods of obtaining the signal model. In addition, we noted that this error can be reduced by transforming both the signal and model to another wavelet domain where Wiener filter can be designed with smaller error.

The main contributions of this paper are: (1) Interpretation of hard and soft thresholding denoising schemes as Wiener filters, (2) Analysis and understanding of *WienerChop* algorithm of [5], and (3) Analysis of the errors in empirical Wiener filters, leading to proper design methods of wavelet domain Wiener filters.

The difficulty in analysis of the wavelet domain Wiener filtering 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 as future research work.

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

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: Change of 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

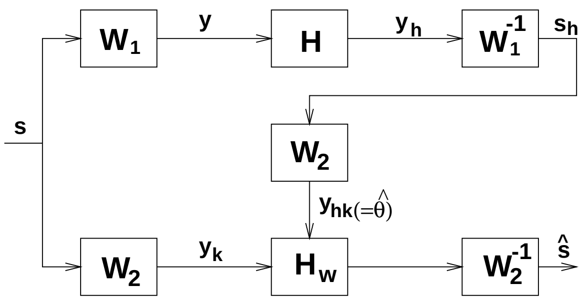

Figure 1: Wavelet-based empirical Wiener filtering. In the upper path, wavelet transform $\mathbf{W_1}$ is used to produce a signal model $\mathbf{y_{hk}}(=\hat{\theta})$. This model is then used to design an empirical Wiener filter $\mathbf{H_w}$ that is applied to the original noisy signal in the $\mathbf{W_2}$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

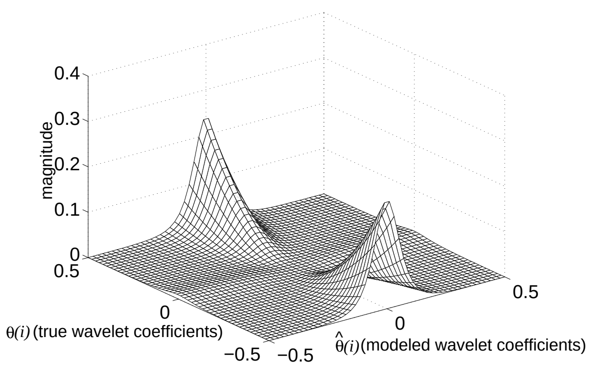

Figure 2: Error due to model mismatch.

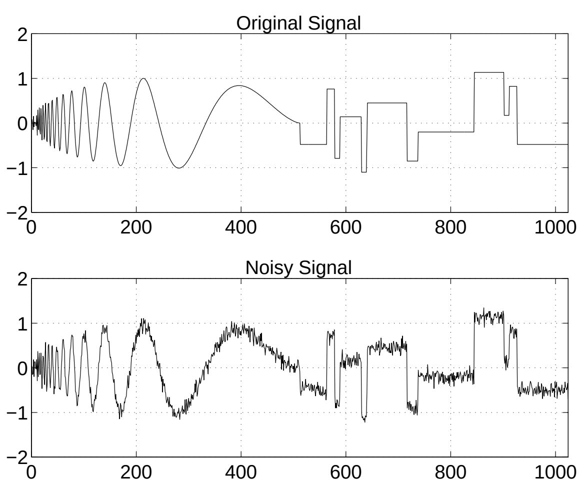

Figure 3: Above: DoppleBlock test signal. Below: signal corrupted by WGN with $\sigma = 0.1$.

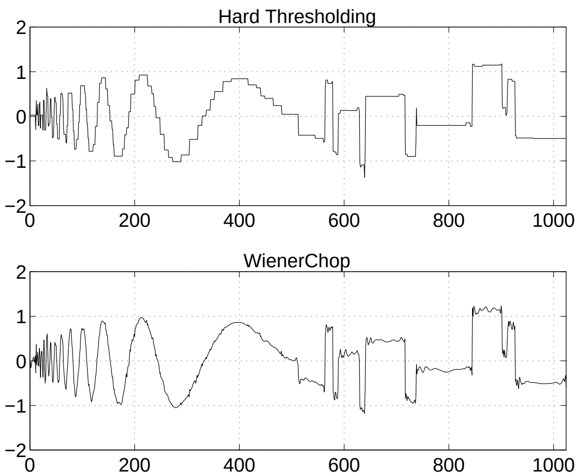

Figure 4: Denoised signals - Above: hard thresholding in D_2, MSE=5.6608, Below: WienerChop using $\mathbf{W}_1 = D_2$ and $\mathbf{W}_2 = D_{12}$, MSE=2.8358.

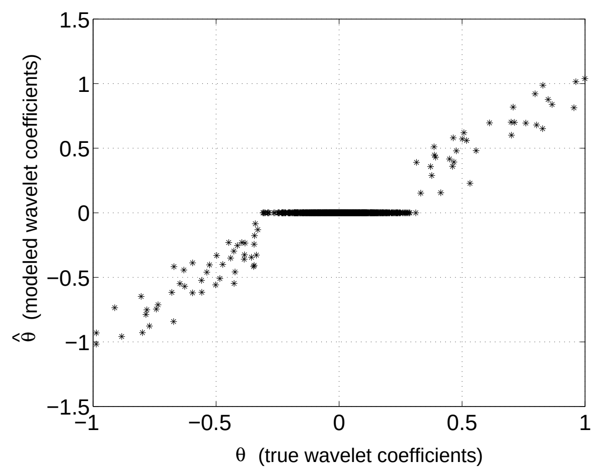

Figure 5: Model mismatch in the $\mathbf{W_1}$ domain.

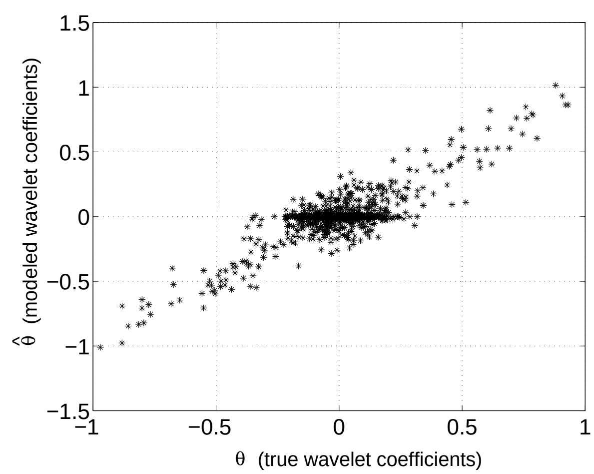

Figure 6: Model mismatch in the $\mathbf{W_2}$ domain.