Improving De-Noising by Coefficient De-Noising and Dyadic Wavelet Transform

Hailong Zhu James T. Kwok
Department of Computer Science
Hong Kong University of Science and Technology
Hong Kong
P.R. China
{zhuhl, jamesk}@cs.ust.hk

LiangSheng Qu
Research Institute of Diagnostics and Cybernetics
Xi'an Jiaotong University
Xi'an
P.R. China
lsqu@xjtu.edu.cn

Abstract

Soft thresholding has been a standard wavelet de-noising procedure in many signal and image processing applications. Theoretically, it is also almost optimal in the sense of nearly achieving the minimax mean-squared error. Inspired by this property, this paper proposes the addition of coefficient de-noising before soft thresholding. This extra step serves to reduce noise in the empirical wavelet coefficients at each scale, and can be shown to yield a lower mean-squared error. Moreover, we advocate the use of the translation-invariant dyadic wavelet transform, together with an approximate self-dual wavelet, instead of the discrete wavelet transform (DWT) in performing de-noising. Experiments show that the proposed method improves the signal-to-noise ratios of the de-noised signals. Moreover, the de-noised signals do not have artifacts typically associated with DWT-based methods.

1. Introduction

Since the past decade, wavelets have been widely used in various signal and image processing tasks, such as for singularity/edge detection, compression and feature extraction [7]. With the inevitable presence of noise in these real-world applications, empirical wavelet coefficients with small magnitudes are likely to contain more noise than information. A simple idea is then to replace the small wavelet coefficients by zero, in the hope of recovering a less noisy signal after performing the inverse wavelet transform. Instead of this *hard thresholding* procedure, Donoho and Johnstone [4] introduced *soft thresholding*, which translates all the empirical wavelet coefficients towards zero. In recent years, several different thresholding procedures have also been investigated (e.g. [1, 5]), though soft thresholding has still remained very competitive. Moreover, soft thresholding enjoys the nice theoretical property that it is almost optimal in the sense of nearly achieving the minimax mean-squared error [3].

In this paper, we investigate on further improving the de-noising performance of soft thresholding by making use of this theoretical property. We show that by preceding soft thresholding by another de-noising procedure (*coefficient de-noising*) that operates on the sequence of wavelet coefficients at each scale, one can obtain a smaller worst-case error than by using soft thresholding alone.

Another issue with de-noising is that it is typically used with the discrete wavelet transform (DWT). However, DWT is not translation-invariant because of the underlying dyadic down-sampling process. Consequently, de-noising procedures based on the DWT often produce undesirable artifacts [2]. In this paper, we address this issue by using instead the dyadic wavelet transform, which constructs a translation-invariant representation by discretizing only the scale parameter, but not the translation parameter. A new wavelet for the dyadic wavelet transform, which has better empirical signal reconstruction performance, is also proposed.

The rest of this paper is organized as follows. Brief introductions on the dyadic wavelet transform and soft thresholding are given in Sections 2 and 3 respectively. Section 4 describes our proposed method. Experimental results are presented in Section 5, and the last section gives some concluding remarks.

2. Dyadic Wavelet Transform

The wavelet transform represents any arbitrary function $f \in L^2(\Re)$ as a superposition of wavelets, which are functions generated from a mother wavelet ψ by dilations b and translations a, as $\psi_{a,b}(x) = \frac{1}{a}\psi(\frac{x-b}{a})$. The wavelet transform Wf of a function f can be written as a convolution product $Wf(a,b) = f \star \bar{\psi}_a(b)$, where $\bar{\psi}_a(b) = \frac{1}{a}\psi^*(\frac{-b}{a})$ and ψ^* is the complex conjugate of ψ.

For computationally efficiency, DWT, which discretizes both the scale and translation parameters along a dyadic sequence, is commonly used. However, it is well-known that the underlying dyadic down-sampling process renders DWT translation variant. The dyadic wavelet transform, on the other hand, constructs a translation-invariant representation by discretizing only the scale[1] but not the translation. For a given ψ, the corresponding reconstructing wavelet (*dual*) χ satisfies the condition

$$S_{\psi,\chi}(\omega) = \sum_{j=-\infty}^{\infty} \hat{\psi}(2^j\omega)\hat{\chi}(2^j\omega) = 1 \qquad (1)$$

almost everywhere, where $\hat{\psi}$ and $\hat{\chi}$ denote the Fourier transforms of ψ and χ respectively. The original function f can then be perfectly reconstructed from Wf as $\sum_{j=-\infty}^{\infty} \frac{1}{2^j} Wf(2^j,\cdot) \star \chi_{2^j}(x)$ [7].

3. Soft Thresholding

After performing the (forward) wavelet transform on the observed signal, a set of noisy empirical wavelet coefficients $\{y_i\}$ are obtained at each scale. Assume that these y_i's follow the model[2] [3]:

$$y_i = \theta_i + \delta u_i, \quad i \in I, \qquad (2)$$

where θ_i's are the wavelet coefficients of the clean signal, I is the index set for the different translations at this scale, $\delta > 0$ is the noise level and u_i's are nuisance terms satisfying $|u_i| \leq 1$. At a given threshold δ, soft thresholding replaces y_i's that are smaller than δ by zero, while the rest are shrunk in absolute value. In other words, it produces the estimate $\hat{\theta}_i^{(\delta)} = \mathrm{sgn}(y_i)(|y_i| - \delta)_+$. This can be shown to satisfy the *uniform shrinkage* condition $|\hat{\theta}_i^{(\delta)}| \leq |\theta_i|$, $\forall i \in I$.

[1] As in DWT, the scale in the dyadic wavelet transform is sampled along the dyadic sequence $\{2^j\}_{j\in Z}$, where Z is the set of integers.

[2] In this Section, we only describe the simpler case for deterministic noise. The case for random noise model with Gaussian white noise $N(0,1)$ is analogous [3].

Besides, if the worst-case error for an estimate $\hat{\theta}$ is defined as $E_\delta(\hat{\theta},\theta) = \sup_{|u_i|\leq 1} \|\hat{\theta} - \theta\|_2^2$, then

$$E_\delta(\hat{\theta}^{(\delta)},\theta) = \sum_{i\in I} \min(\theta_i^2, 4\delta^2) \qquad (3)$$

is almost equal to the minimax error $\inf_{\hat{\theta}} \sup_\Theta E_\delta(\hat{\theta},\theta)$, where Θ is a set of possible θ's. Moreover, for any $\hat{\theta}$ that satisfies the uniform shrinkage condition, $E_\delta(\hat{\theta},\theta) \geq E_\delta(\hat{\theta}^{(\delta)},\theta)$ [3].

4. Proposed Method

4.1. Approximate Self-Dual Wavelet

In [6, 8], the following wavelet (denoted ψ_A hereafter), with an integer parameter p, satisfying

$$\hat{\psi}_A(2^j\omega) = -\frac{(2^j\omega)^2}{4}\left(\frac{\sin(2^j\omega/4)}{2^j\omega/4}\right)^{2p+2},$$

is used for the dyadic wavelet transform. In this paper, we will use another wavelet (denoted ψ_B) which satisfies

$$\hat{\psi}_B(2^j\omega) = 2^j\omega\exp(-(2^j\omega)^2/2).$$

This wavelet has been used in the WaveLab package[3] for performing the continuous wavelet transform. Its dual is difficult to obtain. However, if we use ψ_B as its own dual, condition (1) can be satisfied almost everywhere with very high numerical precision (i.e. $S_{\psi_B,\psi_B}(\omega)$ is very close to 1). Hence, ψ_B can be considered as having the *approximate self-dual* property.

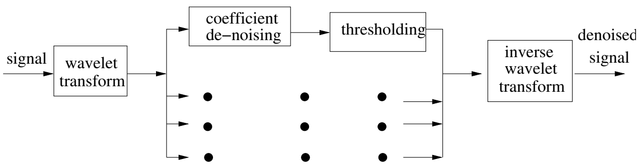
Using a set of 20 signals from the WaveLab, we have experimentally compared the reconstruction performance of the dyadic wavelet transform using ψ_A and ψ_B. The signal-to-noise ratios (SNR's) obtained with ψ_B are much higher than those with ψ_A. However, detailed results will not be reported here because of the lack of space.

4.2. Coefficient De-noising

Recall that soft thresholding yields the smallest worst-case error $E_\delta(\hat{\theta}^{(\delta)},\theta)$ among all estimators that satisfy the uniform shrinkage condition (Section 3). We also notice from (3) that $E_\delta(\hat{\theta}^{(\delta)},\theta)$ is related to the noise level δ. Hence, by reducing δ, we might be able to obtain a smaller $E_\delta(\hat{\theta}^{(\delta)},\theta)$ and consequently improved de-noising performance. To obtain such a reduction, we here propose treating the coefficient sequence $\{y_i\}$ at each scale as a signal and precede wavelet thresholding by another de-noising

[3] The WaveLab package can be obtained from http://www-stat.stanford.edu/~wavelab.

procedure (*coefficient de-noising*). Figure 1 shows the proposed scheme. Notice that it reduces to traditional signal de-noising if coefficient de-noising is removed (and DWT is used as the wavelet transform). In principle, any de-noising procedure can be used for coefficient de-noising. In this paper, we first perform DWT on the (dyadic) wavelet coefficient sequence, then use HYBRID from the WaveLab for thresholding the DWT coefficients, and finally recover the de-noised (dyadic wavelet transform) coefficients by performing the inverse DWT.



(different rows correspond to different scales)

Figure 1. The proposed method.

Assume that the set of de-noised wavelet coefficients $\{y_i'\}$ still follow a similar model as in (2): $y_i' = \theta_i + \delta' u_i'$, where $|u_i'| \leq 1$. Typically, the SNR of $\{y_i'\}$ is larger than that of $\{y_i\}$[4], and thus we may assume that $\delta' \leq \delta$. Applying soft thresholding on $\{y_i'\}$, we obtain from (3) that $E_{\delta'}(\hat{\theta}^{(\delta')}, \theta) = \sum_{i \in I} \min(\theta_i^2, 4\delta'^2) \leq E_\delta(\hat{\theta}^{(\delta)}, \theta)$. In general, the choice of δ' depends on the coefficient de-noising procedure. Here, we employ the simpler approach of using the δ in (2). Following the analysis in [3], one can show that $E_{\delta'}(\hat{\theta}^{(\delta)}, \theta) = \sum_{i \in I} \min(\theta_i^2, (\delta + \delta')^2) \leq E_\delta(\hat{\theta}^{(\delta)}, \theta)$, and thus a smaller worst-case error can still be obtained.

Notice that different δ''s have to be used at different scales. In this paper, we take a sampling approach for the estimation of δ's. As we typically assume Gaussian white noise $N(0, 1)$ [3], we generate a set of pure noise sequences by repeated sampling from $N(0, 1)$. Dyadic wavelet transform is then performed on this set. For each scale, we choose the 95% percentile of the absolute value of the coefficients as the δ.

Moreover, in practice, the performance of de-noising procedures typically deteriorates when the signal is short. With the dyadic wavelet transform, the numbers of wavelet coefficients at all scales are equal to the signal length. On the other hand, with the DWT, the numbers of wavelet coefficients at the coarse scales will be very small (equal to $1, 2, 4, 8, \ldots$). This offers another justification for using the dyadic wavelet transform instead of the DWT.

[4]This assumption is reasonable if the de-noising procedure is reasonably good and the original SNR of $\{y_i\}$ is not very high.

5. Experiments

Because of the lack of space, we will only report experiments on the "Blocks" signal from the WaveLab. The signal is sampled on $[0, 1]$, with a sample size of 1024. We compare our proposed method (DYD-CD) with other signal de-noising routines, including the standard ones of VISU, SURE, HYBRID and MINIMAX from the WaveLab, together with the recently proposed method VM-W based on the VC-theory [1], and the translation-invariant de-noising procedure from [2] (denoted DWT-TI here)[5]. All these methods are used with their default settings. Moreover, notice that all except ours perform DWT on the signal.

Figure 2 compares the various de-noised signals for a particular realization of the noise, at a source SNR of 16dB. The de-noised signal obtained with DYD-CD is closest to the original signal and also does not have artifacts that are associated with the other DWT-based methods.

Figure 3 shows the boxplot for the SNR distributions of the de-noised signals. The results are based on 300 random realizations, at source SNR's of 16dB and 2dB respectively. DYD-CD clearly outperforms the others.

6. Conclusion

In this paper, we propose the addition of coefficient de-noising before wavelet thresholding. This serves to lower the noise level of the wavelet coefficients and can be shown to lead to a lower worst-case error. Moreover, we advocate the use of the translation-invariant dyadic wavelet transform, together with an approximate self-dual wavelet, instead of DWT in performing signal de-noising. Experiments show that the de-noised signals obtained with this proposed method have improved SNR's. Moreover, they do not have artifacts typically resulted from the non-stationary nature of DWT. A disadvantage with using the dyadic wavelet transform, however, is that it is a redundant transform and thus is more expensive than DWT in terms of both storage and time.

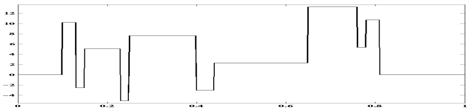
In the future, we will extend this method for image processing, which can be done by simply replacing all 1-D transforms by 2-D transforms.

References

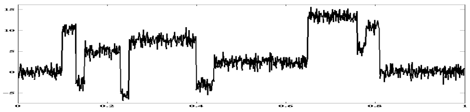
[1] V. Cherkassky and X. Shao. Signal estimation and denoising using VC-theory. *Neural Networks*, 14:37–52, 2001.

[2] R. Coifman and D. Donoho. Translation-invariant de-noising. In A. Antoniadis and G. Oppenheim, editors, *Wavelets in Statistics*, volume 103 of *Lecture Notes in Statistics*, pages 125–150. Springer Verlag, New York, 1995.

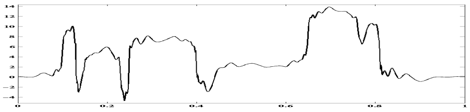
[5]Here, HYBRID is used as the underlying thresholding procedure of DWT-TI.



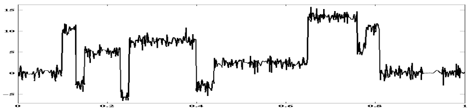
(a) Original signal.



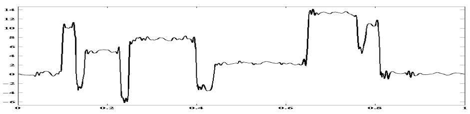
(b) Corrupted signal.



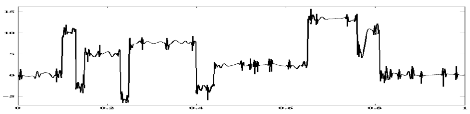
(c) VISU.



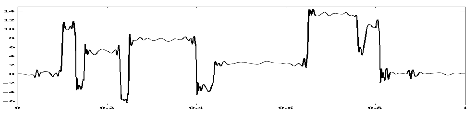
(d) SURE.



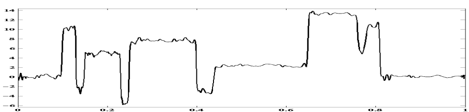
(e) HYBRID.



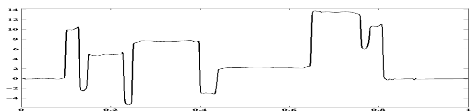
(f) MINIMAX.



(g) VM-W.

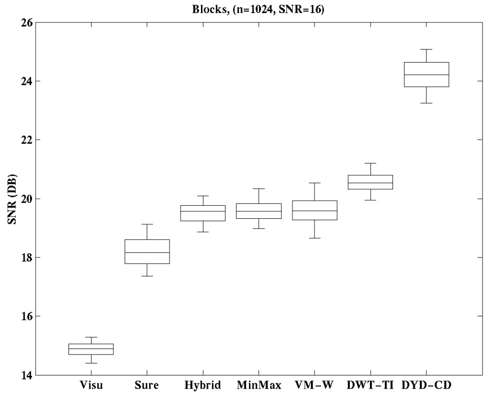


(h) DWT-TI.

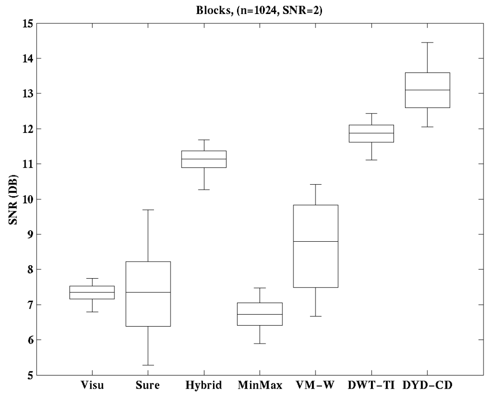


(i) DYD-CD.

Figure 2. De-noised signals at SNR=16dB.



(a) SNR=16dB.



(b) SNR=2dB.

Figure 3. SNR's of the de-noised signals.

[3] D. Donoho. De-noising by soft-thresholding. *IEEE Transactions on Information Theory*, 41(3):613–627, May 1995.

[4] D. Donoho and I. Johnstone. Ideal spatial adaptation via wavelet shrinkage. *Biometrica*, 81:425–455, 1994.

[5] M. Jansen, M. Malfait, and A. Bultheel. Generalized cross validation for wavelet thresholding. *Signal Processing*, 56(1):33–44, Jan. 1997.

[6] A. Laine, J. Fan, and S. Schuler. A framework for contrast enhancement by dyadic wavelet analysis. In A. Gale, S. Astley, D. Dance, and A. Cairns, editors, *Digital Mammography*, pages 91–100. Elsevier, Amsterdam, The Netherlands, 1994.

[7] S. Mallat. *A Wavelet Tour of Signal Processing*. Academic Press, San Diego, 2nd edition, 1999.

[8] S. Mallat and S. Zhong. Characterization of signals from multiscale edges. *IEEE Transactions on Pattern Analysis and Machine Learning*, 7(14):710–732, July 1992.