A NEW TIME-SCALE ADAPTIVE DENOISING METHOD BASED ON WAVELET SHRINKAGE†

Xiao-Ping Zhang and Zhi-Quan Luo

Communications Research Laboratory
McMaster University

ABSTRACT

The wavelet shrinkage denoising approach is able to maintain local regularity of a signal while suppressing noise. However, the conventional wavelet shrinkage based methods are not time-scale adaptive to track the local time-scale variation. In this paper, a new time-scale adaptive denoising method for deterministic signal estimation is presented, based on the wavelet shrinkage. A class of smooth shrinkage functions and the local *SURE* (Stein's Unbiased Risk Estimate) risk are employed to achieve time-scale adaptive denoising. The system structure and the learning algorithm are developed. The numerical results of our system are presented and compared with the conventional wavelet shrinkage techniques as well as their optimal solutions. Results indicate that the new time-scale adaptive method is superior to the conventional methods. It is also shown that the new method sometimes even achieves better performance than the optimal solution of the conventional wavelet shrinkage techniques.

1. INTRODUCTION

Denoising a given noise corrupted signal is a traditional problem in both statistics and in signal processing applications. Linear denoising methods are not so effective when transient nonstationary wideband components are involved since they have similar spectrum to the noise [1]. Recently, Donoho *et al.* [2,3] developed a nonlinear wavelet shrinkage denoising method for statistical applications. The wavelet shrinkage methods rely on the basic idea that the energy of a signal (with some smoothness) will often be concentrated in a few coefficients in wavelet domain while the energy of noise is spread among all coefficients in wavelet domain. Therefore, the nonlinear shrinkage function in wavelet domain will tend to keep a few larger coefficients representing the signal while the noise coefficients will tend to reduce to zero. Normally, the soft-thresholding function, $\eta_s(x,t) = \mathrm{sgn}(x)(|x|-t)_+$, is used as the standard soft-thresholding function. The wavelet shrinkage methods achieve asymptotically near optimal minimax mean-square error for a wide range of signals corrupted by additive white Gaussian noise and retain a good visual effect [2,3]. However, conventional wavelet shrinkage methods are not time-scale adaptive and do not have the capability to track the local time-scale variation in signals. All the parameters are independent on time and often preset. This is partially because the standard shrinkage function does not have high-order derivatives and thus the gradient based adaptive schemes are not tangible in many cases. Recently, a new type of shrinkage functions has been developed by Zhang *et al.* [1] and are defined as follows:

$$\eta_k(x,t) = \begin{cases} x + t - \frac{t}{2k+1}, & x < -t \\ \frac{1}{(2k+1)t^{2k}} \cdot x^{2k+1}, & |x| \le t \\ x - t + \frac{t}{2k+1}, & x > t. \end{cases} \quad (1)$$

They have similar shrinkage properties to the standard shrinkage function and have been proved to have better numerical properties in some applications. (These functions are illustrated in Fig. 1.) Unlike the standard shrinkage function, these shrinkage functions have high-order derivatives so that to develop the gradient based adaptive schemes in wavelet shrinkage methods become tangible.

In this paper, a new time-scale adaptive wavelet shrinkage denoising method for deterministic signal estimation is presented. The new system employs the local *SURE* (Stein's Unbiased Risk Estimate) risk and scale dependent wavelet shrinkage method. The gradient based adaptive algorithm to find time-scale adaptive thresholds is developed by using the shrinkage functions in Eq. (1). We take advantage of the local time-scale information of both the signal and the noise in the new method. It is fully adaptive with respect to the time and scale. Numerical simulations are presented using the new method and the results are compared with conventional wavelet shrinkage methods. It is proved that the performance of the new method is much better than the conventional wavelet shrinkage methods in the MSE

† In *Proc. of ICASSP'99*, Phoenix, Arizona, Mar. 15-19, 1999.
Contact: Dr. X.-P. Zhang, xpzhang@ieee.org, http://www.ee.ryerson.ca/~xzhang

sense. Moreover, it is also shown that the new method often performs better than the optimal solution of the conventional wavelet shrinkage methods.

2. TIME-SCALE ADAPTIVE DENOISING USING WAVELET SHRINKAGE

Assume that the observed data vector $\boldsymbol{y} = [y_0, y_1, \cdots y_{N-1}]^T$ is given by $y_i = f_i + n_i$, $i = 0,1,\ldots,N-1$, where f_i is samples of a deterministic signal f and n is Gaussian white noise with i.i.d. distribution $N(0,\sigma)$. The denoising objective is to estimate real signal f from the observed data vector $\boldsymbol{y}$ with minimum mean square error (MSE), i.e., to minimize MSE error of the estimation $\hat{\boldsymbol{f}}$ from a given noise corrupted signal $\boldsymbol{y}$:

$$R(\hat{\boldsymbol{f}}, \boldsymbol{f}) = \frac{1}{N}\left(\hat{\boldsymbol{f}} - \boldsymbol{f}\right)^2 = \frac{1}{N}\sum_i \left(\hat{f}_i - f_i\right)^2 . \qquad (2)$$

Here we use the mean instead of the mathematical expectation because the optimal solution is desired for each observed data vector $\boldsymbol{y}$. In our new time denoising method, the orthogonal discrete wavelet transform (DWT) is used. The risk function given in (2) can be expressed in wavelet domain:

$$R(\hat{\boldsymbol{f}}, \boldsymbol{f}) = \frac{1}{N}\left(\hat{\boldsymbol{f}} - \boldsymbol{f}\right)^2 = \frac{1}{N}\sum_{j,k} \left(\hat{v}(j,m) - v(j,m)\right)^2 , \quad (3)$$

where $v(j,m)$ represents the m-th wavelet coefficient of the signal $\boldsymbol{f}$ at scale j, $j = 1,...,J$. Note that only the scaling coefficients at the largest scale J are useful in DWT and $v(0,m)$ is used to represent the m-th scaling coefficient at the largest scale J. Similarly, $\hat{v}(j,m)$ and $u(j,m)$ represent the wavelet coefficients of the estimate $\hat{\boldsymbol{f}}$ and the observed data vector $\boldsymbol{y}$, respectively.

To achieve the time-scale adaptive denoising, we need to find the time series of the signal in wavelet domain by rearranging the wavelet coefficients. Suppose the signal samples enter the DWT processing block in a time order. Then the time series v_i of the wavelet coefficients can be obtained as

$$\begin{aligned}\boldsymbol{v} &= [v_0, v_1, v_2, v_3, v_4, v_5, v_6, v_7, \cdots, v_{2^J-1}, v_{2^J}, \cdots, v_{N-1}]^T \\ &= [v(1,0), v(1,1), v(2,0), v(1,2), v(1,3), v(2,1), v(3,0), \\ &\qquad \cdots, v(J,0), v(0,0), v(1,2^{J+1}), \cdots, v(0, M-1)]^T .\end{aligned}$$

That means, at the time we have two wavelet coefficients at scale j, we can obtain one wavelet (and scaling) coefficient at the scale $j+1$ due to the downsampling operation in DWT. Here M is the number of the scaling coefficients at the scale J. The wavelet coefficients time series v_i and u_i can also be constructed in the same way. Then Eq. (3) becomes

$$R(\hat{\boldsymbol{f}}, \boldsymbol{f}) = \frac{1}{N}\sum_i \left(\hat{v}_i - v_i\right)^2 . \qquad (4)$$

We note that the conventional scale dependent shrinkage method is not fully scale adaptive since the thresholds at each scale are not selected using a fully adaptive algorithm. Since it has been proved that the scale dependent shrinkage scheme has better performance than universal shrinkage scheme [2-4], our new method is developed based on scale dependent shrinkage scheme, i.e., different thresholds will be used for the wavelet coefficients at different scales.

In the new method, the estimated wavelet coefficients $\hat{v}$ are obtained using time-scale adaptive shrinkage of wavelet coefficients u of observed data y, i.e., $\hat{v}(j,m) = \eta_k(u(j,m), t(j,m))$. For a time series u_i, it can be written as $\hat{v}_i = \eta_k(u_i, \boldsymbol{t}(i))$, $\boldsymbol{t}(i)$ denotes vector $[t_1(i), t_2(i), \ldots, t_J(i)]^T$, $t_j(i)$ is the time-scale adaptive threshold at scale j and time i. Note that the scaling coefficients $u(0,m)$ are not shrunk since they contain the basic information of the signal and only one $t_j(i)$ is employed to perform shrinkage on u_i at each time i. In our nonlinear time-scale adaptive denoising, we will attempt to select the time-scale adaptive parameter $t_j(i)$ in the nonlinear shrinkage function $\eta_k(x,t)$ to minimize the MSE risk.

The MSE risk Eq. (4) is not known since the real signal f is not known. We will estimate the MSE risk based on the *SURE* risk. The *SURE* risk is established for signal estimation [3,7] as follows: assume $\boldsymbol{g}(\boldsymbol{u}, \boldsymbol{t}) = \hat{\boldsymbol{v}} - \boldsymbol{u}$, where $\boldsymbol{g} = [g_0, g_1, \ldots, g_{N-1}]^T$ is a mapping from $\boldsymbol{R}^N$ to $\boldsymbol{R}^N$ and $\boldsymbol{t}$ denotes vector $[t_1, t_2, \ldots, t_J]^T$ in scale dependent shrinkage scheme. The *Stein's Unbiased Risk Estimate* (*SURE*) is defined as:

$$R_s(\boldsymbol{t}) = N + \left\|\boldsymbol{g}(\boldsymbol{u})\right\|^2 + 2\nabla_y \cdot \boldsymbol{g}(\boldsymbol{u}) , \qquad (5)$$

where

$$\nabla_y \cdot \boldsymbol{g}(\boldsymbol{u}) = \sum_{i=0}^{N} \frac{\partial g_i}{\partial u_i} \qquad (6)$$

and

$$\partial g_i / \partial u_i = \partial \eta_k(x, \boldsymbol{t}) / \partial x \Big|_{x = u_i} - 1 . \qquad (7)$$

Stein [7] showed that when $\boldsymbol{g}(\boldsymbol{u})$ is weakly differentiable, the *SURE* risk is an unbiased estimator of the MSE risk. However, to obtain the local time-adaptive parameters, the local MSE risk needs to be estimated. Therefore, we will use the local *SURE* risk at each time i:

$$R_s^i(t) = 1 + \left\|g(u_i)\right\|^2 + 2\nabla_y \cdot g(u_i) . \qquad (8)$$

Now, the time-scale adaptive denoising algorithm based on $R_s^i(t)$ can be developed as follows to minimize the MSE risk by adjusting parameter $\boldsymbol{t}$ at each time.

Step 1. Initialize parameter $\boldsymbol{t}^0(0)=\boldsymbol{t}_0$, $l=0$.

Step 2. At learning iteration l, for each input time sample u_i, $i=0,\ldots,N-1$, in wavelet domain, adjust $\boldsymbol{t}(i)$ using following scheme,

$$\boldsymbol{t}^l(i+1) = \boldsymbol{t}^l(i) - \Delta\boldsymbol{t}^l(i)\ ,\ i=0,\ldots,N-1, \tag{9}$$

where
$$\Delta\boldsymbol{t}^l(i) = \boldsymbol{\alpha}(i)\cdot\left.\frac{\partial R_s^i(\boldsymbol{t})}{\partial \boldsymbol{t}}\right|_{\boldsymbol{t}=\boldsymbol{t}^l(i)}, \tag{10}$$

where $\boldsymbol{\alpha}(i) = diag[\alpha_1(i), \alpha_2(i), \cdots, \alpha_J(i)]$ is the training rate matrix of each step and $\alpha_j(i)$ is the learning rate for parameter t_j.

Step 3. Set $l=l+1$ and $\boldsymbol{t}^l(0)=\boldsymbol{t}^{l-1}(N-1)$. Repeat step 2 and 3 for u_i, $i=0,\ldots,N-1$, until certain convergence criterion is satisfied or the maximum training times are reached.

From Eq. (7) and (10), it is apparent that to calculate the gradient of the *SURE* risk R_s^i, the second derivatives of the shrinkage function have to be employed. Note that although the standard soft-thresholding function $\eta_s(x,t)$ is weakly differentiable in Stein's sense [7], it does not have the second derivatives. Therefore, it is not possible to use the standard soft-thresholding function to achieve the above time-scale adaptive denoising scheme. Therefore, here we employ the shrinkage function $\eta_k(x,t)$ of Eq. (1) and then the gradients of R_s^i can be calculated accordingly.

In this way, the time-scale adaptive parameter $\boldsymbol{t}(i)$, $i=0,\ldots,N-1$, can be obtained based on the local *SURE* risk. The adaptive system structure of the proposed method is shown in Fig. 2. In Fig. 2, y_i, $i=0,\ldots,N-1$, are the input samples. The DWT block is fast discrete wavelet transform block [5,6] and IDWT block is the fast inverse discrete wavelet transform. Samples u_i, $i=0,\ldots,N-1$, are the time series of wavelet coefficients of the input signal in wavelet domain. The time series $\hat{v}_i$, i.e., the estimation of the wavelet coefficients time series v_i of the real signal, are obtained using the nonlinear shrinkage function. The local *SURE* risk R_s^i is used to estimate the MSE and then adjust the local time-scale adaptive thresholds $\boldsymbol{t}(i)$ at each time i by the preceding learning algorithm.

3. NUMERICAL EXAMPLES

The test signal is a Doppler signal, which is used by most of other wavelet shrinkage related literature, as shown in Fig. 3. The data length N is 2048 samples. The input noise corrupted signals of different signal-to-noise-ratio (SNR) are tested using some conventional wavelet shrinkage methods as well as our new time-scale adaptive wavelet shrinkage denoising method. The performance comparison is shown in Table 1. For different input SNR (SNR_{in}) using different wavelet shrinkage methods of estimation, the output SNR (SNR_{out}) are given in Table 1. TS-SURE represents our new time-scale adaptive denoising method using $\eta_3(x,t)$. The initial value of the adaptive method is selected as $t_j^0(0) = \sqrt{2\log N/N}$, i.e., we start with a generally used universal threshold. Convergence criterion for our time-scale adaptive denoising method is $\max_i(t^l(i)-t^{l-1}(i))<10^{-6}$.

For comparison, two typical conventional wavelet shrinkage methods are applied in the numerical examples. *WaveShrink* is the universal thresholding scheme proposed by Donoho [2]. *SUREShrink* is an optimized hybrid scale dependent thresholding scheme based on *SURE* risk [3] which shows the best MSE performance among the conventional wavelet shrinkage denoising methods. The optimal solutions under MSE risk of Eq. (2) are also calculated for universal and scale dependent wavelet shrinkage schemes. UOPT represents the numerical optimal scale dependent threshold selection using standard soft-thresholding function $\eta_s(x,t)$. SOPT represents the numerical optimal scale dependent threshold selection using the standard soft-thresholding function $\eta_s(x,t)$. The Daubechies8 least asymmetrical wavelet [6] is used and the largest DWT level $J=6$ is selected for all above methods.

As indicated in Table 1, our new time-scale adaptive denoising method consistently outperforms the conventional wavelet shrinkage methods (*WaveShrink* and *SUREShrink*) in terms of the output SNR. The new method often performs even better than the optimal solutions (UOPT and SOPT) of the conventional methods. This is likely due to the fact that the threshold in conventional is independent on time. Also, in numerical results, we found that the new method performs much better than the conventional wavelet shrinkage methods when the local signal frequency changes rapidly, i.e., the new method extracts more signal information in the left part of Doppler signal (see Fig. 2) in the example. This indicates that the new time-scale adaptive denoising method does have the capability to track the local time-scale variation in signals. Other numerical simulations show the similar results.

In the new time-scale adaptive denoising method, the adaptive thresholds are selected automatically with respect to both time and scale. This will be very useful in practical applications.

4. CONCLUSION

In this paper, a new type of nonlinear time-scale adaptive denoising system based on wavelet shrinkage scheme has been presented. The new time-scale adaptive algorithm is based on *SURE* risk and certain type of shrinkage functions with high order derivatives. Numerical results show that the new time-scale adaptive denoising method performs much better than conventional wavelet shrinkage methods in MSE sense. In some occasions, the new method even performs better than the optimal solution of the conventional wavelet shrinkage methods. Unlike conventional wavelet shrinkage denoising method, the new method can find the local time-scale adaptive thresholds from the input data samples automatically. Therefore, the new time-scale adaptive denoising method is more practical and effective.

5. REFERENCES

[1] Xiao-Ping Zhang and M. Desai, "Nonlinear adaptive noise suppression based on wavelet transform," in *Proc. ICASSP98*, vol. 3, pp. 1589-1592, Seattle, May 12-15, 1998.

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

[3] D. L. Donoho and I. M. Johnstone, "Adapting to unknown smoothness via wavelet shrinkage," *J. Am. Stat. Ass.*, vol. 90, no. 432, pp. 1200-1224, 1995.

[4] M. Lang, H. Guo, J. E. Odegard, C. S. Burrus, and R. O. Wells, "Noise reduction using an undecimated discrete wavelet transform," *IEEE Signal Processing Letters*, vol. 3, no. 1, pp. 10-12, 1996.

[5] O. Rioul and P. Duhamel, "Fast algorithms for discrete and continuous wavelet transforms", *IEEE Trans. on Info. Theo.*, vol. 38, no. 2, Mar. 1992.

[6] I. Daubechies, *Ten Lectures on Wavelets*, SIAM, Philadelphia, PA, 1992.

[7] C. Stein, "Estimation of the mean of a multivariate normal distribution," *Annals of Statistics*, vol. 9, no. 6, pp. 1135-1151, 1981.

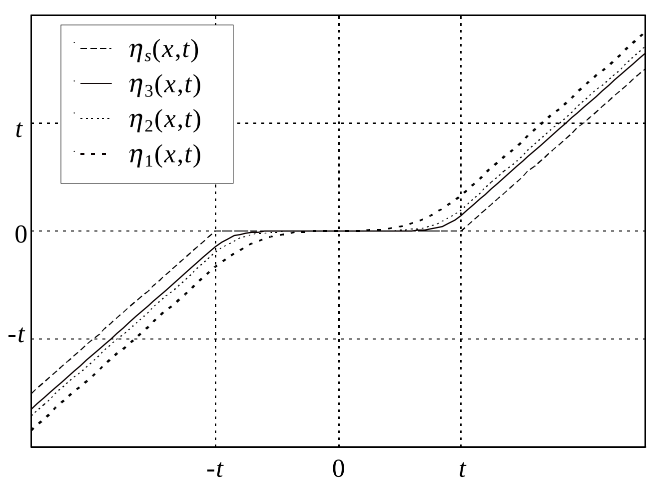

Fig. 1. The standard shrinkage function $\eta_s(x,t)$ (dashed line) and the shrinkage functions in Eq. (1) with k=1,2,3.

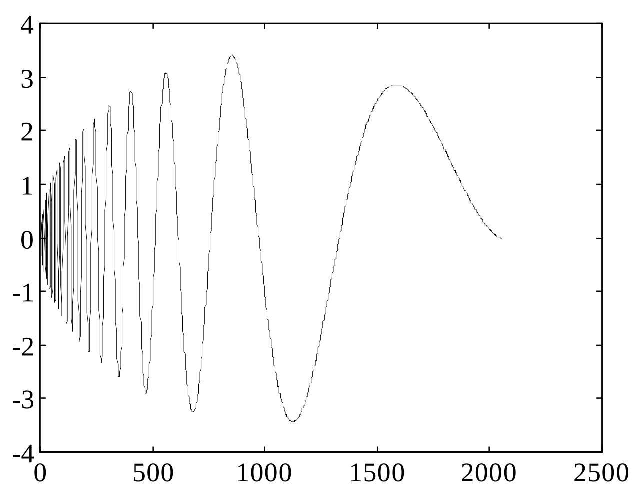

Fig. 3. Test signal (Doppler)

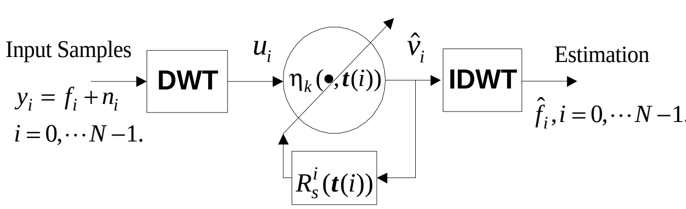

Fig. 2. The time-scale adaptive denoising system

SNR_{in} (dB)	-6	-3	0	3	6
TS-SURE	8.99	9.79	12.06	16.59	18.78
WaveShrink	7.40	8.11	10.14	11.87	14.44
SUREShrink	8.06	9.04	11.98	14.52	16.93
UOPT	7.52	8.44	11.12	14.64	16.95
SOPT	8.92	10.25	12.67	16.37	18.77

Table. 1. The output SNR (dB) of different denoising methods