

Adaptive Denoising Based On SURE Risk

Xiao-Ping Zhang, *Member, IEEE* and M. Desai

Published in IEEE Signal Processing Letters, vol. 5, no. 10, October 1998.

Abstract

In this paper, a new adaptive denoising method is presented based on *Stein's Unbiased Risk Estimate* (SURE) and on a new class of thresholding functions. First we present a new class of thresholding functions which has continuous derivative while the derivative of standard soft-thresholding function is not continuous. The new thresholding functions make it possible to construct the adaptive algorithm whenever using wavelet shrinkage method. By using the new thresholding functions, a new adaptive denoising method is presented based on SURE. Several numerical examples are given. The results indicated that for denoising applications, the proposed method is very effective in adaptively finding the optimal solution in mean square error (MSE) sense. It is also shown that this method gives better MSE performance than those conventional wavelet shrinkage methods.

X.-P. Zhang was with the University of Texas, San Antonio. He is currently with Department of Electrical & Computer Engineering, Ryerson Polytechnic University, 350 Victoria Street, Toronto, Ontario, CANADA, M5B 2K3. E-mail : xpzhang@ieee.org

This work was supported in part by NASA grant no. NAG10-0155 and Texas Advanced Research Program grant no. ARP 002-1996.

I. INTRODUCTION

For denoising applications with known noisy function, it is often necessary to search for the optimal minimum mean-square error (l_2 risk) estimate using *a priori* information. Recently, Donoho and others have developed wavelet shrinkage methods for denoising for the function estimation applications [1], [2], [3], [4], [5], [6]. Of these wavelet shrinkage methods, *SureShrink* is an optimized hybrid scale dependent thresholding scheme based on SURE (*Stein's Unbiased Risk Estimate* [7]) risk [2]. It combines the universal threshold selecting scheme and the scale dependent adaptive threshold selecting scheme and provides the best estimation results in the sense of l_2 risk. SURE risk is a very good estimation of the true l_2 risk when the true function is not known [7]. However, since the standard soft-thresholding function is weakly differentiable only in the first order, it does not allow for the gradient based optimization method to search for the optimal solution for SURE risk. To this end, Donoho's scheme [2] roughly searches for the optimal threshold within a finite threshold set.

In this paper, first a new class of thresholding functions with continuous derivative is presented. Note that the derivative of standard soft-thresholding function is not continuous. The new thresholding functions do the similar manipulation as the standard soft-thresholding function and they make it possible to search for optimal thresholds using gradient based adaptive algorithms. Next by using the new thresholding functions, a new adaptive denoising method is presented based on SURE risk. Several numerical examples are given.

II. ADAPTIVE DENOISING METHOD

A. Objective

Assume the observed data vector $\mathbf{y} = [y_0, y_1, \dots, y_{N-1}]^T$ given by:

$$y_i = f_i + n_i, \quad i = 0, 1, \dots, N - 1, \quad (1)$$

where f_i is samples of a deterministic function f and n is Gaussian white noise with i.i.d. distribution $N(0, \sigma)$. Our denoising objective is to estimate function f with minimum

mean square error (MSE), i.e., to minimize l_2 risk for a given noisy function:

$$R(\hat{f}, f) = \frac{1}{N} \|\hat{\mathbf{f}} - \mathbf{f}\|^2 = \frac{1}{N} \sum_{i=0}^{N-1} (\hat{f}_i - f_i)^2. \quad (2)$$

Note that here we use mean instead of the mathematical expectation because the optimal solution is desired for each individual noisy function. The conventional wavelet shrinkage methods are shown to be effective in minimax MSE sense and retain a good visual effect [1], [2]. There are three steps in wavelet shrinkage methods: (i) Apply the discrete wavelet transform (DWT) [8] to the vector \mathbf{y} and obtain the empirical wavelet coefficients $u_{j,k}$ at scale j , where $j = 1, \dots, J$. Note $u_{0,k}$ represents the scaling coefficients and will not be shrunk in the next step. (ii) Apply the nonlinearity, such as soft-thresholding function, $\eta_s(x, t_j) = \text{sgn}(x)(|x| - t_j)_+$, to the empirical wavelet coefficients at each scale $j, j = 1, \dots, J$. Then the estimate coefficients $\hat{v}_{j,k}$ are obtained based on the selected threshold $\mathbf{t} = [t_1, t_2, \dots, t_J]^T$. Note t_j is the threshold for wavelet coefficients at scale j . In universal thresholding scheme, the same threshold is used at all scales. In scale dependent schemes such as *SureShrink*, the thresholds at different scales are generally different. (iii) Use the inverse DWT on thresholded wavelet coefficients and obtain the estimate of the function.

The wavelet shrinkage method relies on the basic idea that the energy of a function (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 thresholding function in wavelet domain will tend to keep a few larger coefficients representing the function while the noise coefficients will tend to reduce to zero.

In the wavelet shrinkage methods, the orthogonal DWT is often used. If orthogonal DWT is used, then the risk function given in (2) can be expressed as

$$R(\hat{f}, f) = \frac{1}{N} \sum_{i=0}^{N-1} (\hat{f}_i - f_i)^2 = \frac{1}{N} \sum_{j,k} (\hat{v}_{j,k} - v_{j,k})^2. \quad (3)$$

B. A new class of differentiable thresholding functions

The widely used soft-thresholding function are continuous with discontinuous derivative. However, the continuous derivative or higher order derivatives are often desired for optimization problems. A new class of nonlinear soft-like thresholding functions, motivated

by the differentiable *Sigmoid* function which replaces the undifferentiable hard-limited function in traditional neural network [9], with continuous derivatives are constructed as follows:

$$\eta_k(x, t) = \begin{cases} x + t - \frac{t}{2^{k+1}}, & x < -t \\ \frac{1}{(2^{k+1})t^{2k}} \cdot x^{2k+1}, & |x| \leq t \\ x - t + \frac{t}{2^{k+1}}, & x > t, \end{cases} \quad (4)$$

where k is a positive integer. Note that the limit of $\eta_k(x, t)$ when $k \rightarrow \infty$ is just the commonly used soft-thresholding function $\eta_s(x, t)$. The thresholding functions $\eta_s(x, t)$, $\eta_1(x, t)$ and $\eta_3(x, t)$ are shown in Fig. 1. It can be seen that the new thresholding functions perform the similar operations to the standard soft-thresholding function. Therefore, the similar smoothness property of the estimate using the new thresholding functions can be expected. The new thresholding functions $\eta_k(x, t)$ will have better numerical properties as will be shown in the examples.

C. The Adaptive Denoising Using SURE Risk

The SURE risk is defined for function estimation [2], [7]. For function estimation problem shown as in (1), assume

$$\mathbf{g}(\mathbf{y}) = \hat{f}(\mathbf{y}) - \mathbf{y}, \quad (5)$$

where $\mathbf{g} = [g_0, g_1, \dots, g_{N-1}]^T$ is a function from \mathcal{R}^N to \mathcal{R}^N . Stein [7] showed that when $\mathbf{g}(\mathbf{y})$ is weakly differentiable, then

$$E_f \|\hat{f}(\mathbf{y}) - \mathbf{f}\|^2 = N + E_f \left\{ \|\mathbf{g}(\mathbf{y})\|^2 + 2\nabla_{\mathbf{y}} \cdot \mathbf{g}(\mathbf{y}) \right\}, \quad (6)$$

where $\nabla_{\mathbf{y}} \cdot \mathbf{g}(\mathbf{y}) = \sum_{i=0}^{N-1} \frac{\partial g_i}{\partial y_i}$. The *Stein's Unbiased Risk Estimate* (SURE) is an unbiased estimator of the above risk, defined as

$$R_s(t) = N + \|\mathbf{g}(\mathbf{y})\|^2 + 2\nabla_{\mathbf{y}} \cdot \mathbf{g}(\mathbf{y}). \quad (7)$$

Donoho [2] used the standard soft-thresholding function and selected a threshold t^S in set $\{y_0, y_1, \dots, y_{N-1}\}$,

$$t^S = \arg \min_{t \in \{y_0, y_1, \dots, y_{N-1}\}} R_s(t). \quad (8)$$

Note that t^S is a sub-optimal threshold for the risk R_s because it is selected within a finite set.

Although the standard soft-thresholding function $\eta_s(x, t)$ is weakly differentiable in Stein's sense [7], it does not have the second derivative (see Fig. 1). Therefore, it is not possible to look for the optimal solution for the risk R_s using gradient based adaptive algorithm as the case in neural network using hard-limited function. However, the proposed thresholding function $\eta_k(x, t)$ is twice differentiable in Stein's sense. Then the gradients of R_s can be calculated as follows. From Eq. (7), we obtain

$$\frac{\partial R_s(t)}{\partial t} = 2 \sum_{i=0}^{N-1} g_i \cdot \frac{\partial g_i}{\partial t} + 2 \sum_{i=0}^{N-1} \frac{\partial^2 g_i}{\partial y_i \partial t}. \quad (9)$$

From (5), $g_i = \eta_k(y_i, t) - y_i$. Using (4) and (5), we obtain

$$\frac{\partial g_i}{\partial t} = \frac{\partial \eta_k(y_i, t)}{\partial t} = \begin{cases} 1 - \frac{1}{2k+1}, & y_i < -t \\ -\frac{2k}{2k+1} \cdot \left(\frac{y_i}{t}\right)^{2k+1}, & |y_i| \leq t \\ -1 + \frac{1}{2k+1}, & y_i > t. \end{cases} \quad (10)$$

and

$$\frac{\partial^2 g_i}{\partial y_i \partial t} = \begin{cases} 0, & |y_i| > t \\ -\frac{2k}{t^{2k+1}} \cdot y_i^{2k}, & |y_i| \leq t. \end{cases} \quad (11)$$

Using Eq. (3), the SURE risk $R_s^{(j)}$ at each scale j can be calculated in wavelet domain independently. Then the gradient based learning algorithm can be used to search for the optimal scale dependent threshold in wavelet domain under the estimated risk $R_s^{(j)}$. The threshold can be adjusted toward the optimal solution under SURE risk using following adaptive steps:

$$t_j(i+1) = t_j(i) - \Delta t_j(i), j = 1, 2, \dots, J, \quad (12)$$

where the adjustment of threshold at step i

$$\Delta t_j(i) = \alpha_j(i) \cdot \left. \frac{\partial R_s^{(j)}}{\partial t_j} \right|_{t_j=t_j(i)} \quad (13)$$

in each step, where $\alpha_j(i)$ is the learning rate. Note that $\alpha_j(i)$ should be selected so that $t_j(i+1) \geq 0$.

Since the SURE risk is also an estimate of the risk defined in (2), the optimal solution under SURE risk will be an estimate of the optimal solution for risk (2).

III. NUMERICAL EXAMPLES

The artificial functions such as Blocks, Bumps, HeaviSine and Doppler, created by Donoho *et al* [2] are selected as test functions. Assume the length of all the functions is 2048 with the signal to noise ratio $SNR = \frac{\sigma_f}{\sigma_n} = 7$. The Daubechies8 least asymmetrical wavelet [10] is used and the largest DWT level $J = 6$ is selected. The l_2 risks (MSEs) in Eq. (2) of different wavelet shrinkage methods for estimation are given in Table I. *WaveShrink* is the universal thresholding scheme proposed by Donoho [1]. *SureShrink* is an optimized hybrid scale dependent thresholding scheme based on SURE risk [2]. ADPSURE represents the proposed adaptive denoise method using $\eta_3(x, t)$. The initial value of the adaptive method is selected as $t_j(0) = \sqrt{2 \log N/N}$. Convergence criterion for our optimal threshold selection is $\Delta t(i)/t(i) < 1e - 6$. For comparison, the numerical optimal solutions under risk (2) are also calculated for universal and scale dependent wavelet shrinkage schemes. SOPT0 represents the numerical optimal scale dependent threshold selection using standard soft-thresholding function $\eta_s(x, t)$. SOPT1 represents the numerical optimal scale dependent threshold selection using the proposed thresholding function $\eta_3(x, t)$.

As can be seen in Table I, the optimal solutions using the proposed new thresholding function (SOPT1) are better than the optimal solutions using the standard soft-thresholding function (SOPT0). Note that in Table I that the ADPSURE method performs better for all functions and the results of the ADPSURE method are very close to the optimal solutions.

IV. CONCLUSION

In this paper, a new class of nonlinear thresholding functions were presented. Unlike the standard soft-thresholding function, these functions have continuous derivatives. Since the new thresholding functions perform similar operations to the standard thresholding function, similar smoothness property of the estimate using the new thresholding functions can be expected. The significance of the new thresholding functions is that they make it possible to search for optimal thresholds using gradient based adaptive algorithms. Based on SURE risk, an adaptive wavelet shrinkage method has been presented. The numerical re-

sults show that the proposed method performs better than conventional wavelet shrinkage method. It can be also seen that the optimal solution using the new thresholding function is better than the standard soft-thresholding function for the given examples. The new thresholding functions can also be used for other shrinkage methods such as undecimated wavelet shrinkage methods in [3], [6]. Furthermore, the adaptive wavelet shrinkage method using new thresholding functions can be used in real-time adaptive signal processing [11].

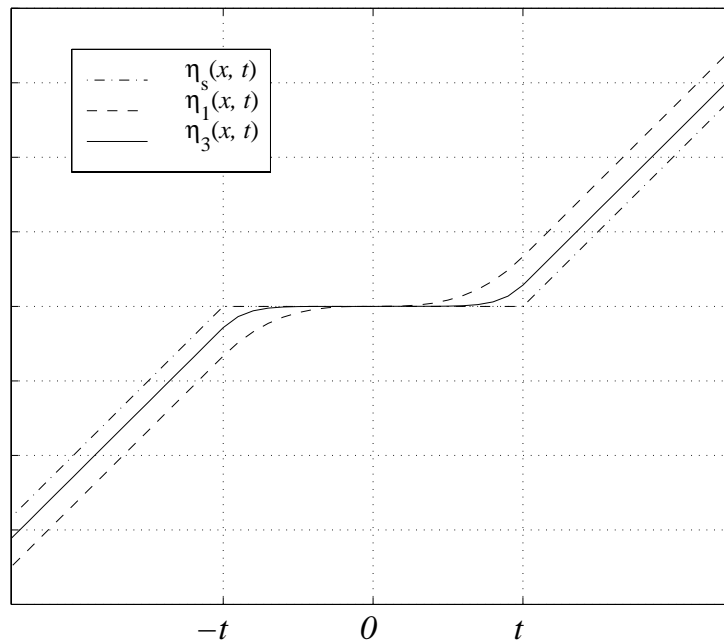
REFERENCES

- [1] D. L. Donoho, "De-noising by soft-thresholding," *IEEE Trans. Inform. Theory*, vol. 41, no. 3, pp. 613–627, May 1995.
- [2] D. L. Donoho and I. M. Johnstone, "Adaptating to unknown smoothness via wavelet shrinkage," *J. Am. Stat. Ass.*, vol. 90, no. 432, pp. 1200–1224, 1995.
- [3] R. R. Coifman and D. L. Donoho, "Translation-invariant de-noising," Tech. Rep., Stanford University, Department of Statistics, 1995, To Appear, *Wavelets and Statistics*, Anestis Antoniadis, ed. Springer-Verlag Lecture Notes.
- [4] G. P. Nason, "Wavelet shrinkage by cross-validation," *Journal of the Royal Statistical Society B*, vol. 58, pp. 463–479, 1996.
- [5] A. Bruce and H.-Y. Gao, "Understanding WaveShrink: Variance and bias estimation," *Biometrika*, vol. 83, no. 4, 1996.
- [6] 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.
- [7] C. Stein, "Estimation of the mean of a multivariate normal distribution," *Annals of Statistics*, vol. 9, no. 6, pp. 1135–1151, 1981.
- [8] O. Rioul and P. Duhamel, "Fast algorithms for discrete and continuous wavelet transforms," *IEEE Trans. on Info. Theo.*, vol. 38, no. 2, pp. 569–586, Mar. 1992.
- [9] S. Haykin, *Neural Network: A Comprehensive Foundation*, Prentice-Hall, Upper Saddle River, New Jersey, 1994.
- [10] I. Daubechies, *Ten Lectures on Wavelets*, SIAM, Philadelphia, PA, 1992.
- [11] X.-P. Zhang and M. Desai, "Nonlinear adaptive noise suppression based on wavelet transform," *submitted to ICASSP'98*, 1997.

	Blocks	Bumps	HeaviSine	Doppler
Original	0.9866	1.0038	1.0591	0.9669
<i>WaveShrink</i>	0.9080	1.0471	0.0952	0.3207
<i>SureShrink</i>	0.3366	0.2846	0.0612	0.1372
ADPSURE	0.2352	0.2559	0.0559	0.1052
SOPT0	0.2255	0.2501	0.0522	0.0934
SOPT1	0.2252	0.2493	0.0484	0.0934

TABLE I

MSE RISK FOR DIFFERENT DENOISING SCHEMES.

Fig. 1. Different thresholding functions $\eta_s(x, t)$, $\eta_1(x, t)$ and $\eta_3(x, t)$.