NONLINEAR ADAPTIVE NOISE SUPPRESSION BASED ON WAVELET TRANSFORM

Xiao-Ping Zhang and Mita D. Desai

Division of Engineering
University of Texas at San Antonio
San Antonio, TX 78249, USA
Email: xiao, desai@runner.utsa.edu

ABSTRACT

The conventional linear adaptive filters are not effective for discriminating the transient wideband signal components from noise. A recently developed wavelet shrinkage approach is able to maintain the function local regularity while suppressing noise however, it has only been used in function estimation problems. In this paper, a new type of nonlinear filtering method for adaptive noise suppression is presented, based on shrinkage method. A new class of shrinkage functions is also presented. The filtering structure and the learning algorithm are developed. The theoretical analysis proves convergence in certain statistical sense. The numerical results of our system are presented for both the standard and the new shrinkage function and compared with the conventional linear adaptive filter based techniques. Results indicate that both the optimal solution and the learning performance are superior to the conventional methods. It is shown that our new shrinkage function performs better than the standard shrinkage function.

1. INTRODUCTION

Large part of the adaptive signal processing deals with the problem of noise suppression [1, 2]. However, the linear filter is a simple modification of the spectrum of the signal because the complex exponential functions $e^{j\omega}$ are the eigenfunctions of any linear system. Therefore, when the signal contains transient wideband components, linear filters are not effective for removing noise. For example, some transient impulses can cause wideband components in the signal which often contain important information. In such cases, linear filters in general are not capable of discriminating nonstationary wideband signal components from noise since both have similar spectrum. Also, several authors such as [3, 4] have investigated wavelet based linear adaptive filtering techniques. These methods show better learning performance than the linear adaptive filters in time domain. However, these techniques are classified as linear systems since wavelet transform (WT) is a linear transform. The optimal solution of these methods is the same as the linear filtering. Therefore, the essential disadvantage of linear filtering methods will still exist.

Recently, Donoho [5, 6], and others have developed wavelet shrinkage methods for statistical applications. The main purpose of the method is to estimate a wide class of functions in some smoothness spaces from their corrupted (by additive Gaussian noise) version [5]. The estimation achieves asymptotically near optimal minimax mean-square error over a wide range of smoothness classes and keeps the regularity of the function at the same time. However, there are some essential differences between function estimation in noise and noise suppression in signals. For function estimation, the object is deterministic function and all of the data samples of the function are used. For adaptive noise suppression in signal processing applications, we assume the signal is time-varying random process and only the past data samples are known. The objective of the adaptive system is to track the changes in the system in real-time and continually seek the optimum in some statistical sense. Furthermore, in signal processing applications, we search for the optimal minimum mean-square error solution using *a priori* information for a specific signal. The optimal minimax solution often only has theoretical meaning because it is often far from the optimal solution for a specific practical problem.

In this paper, first a new class of nonlinear shrinkage functions is presented. Unlike the standard soft-thresholding function, these new nonlinear shrinkage functions have continuous derivative. Then a new nonlinear filtering system using wavelet shrinkage method is presented for adaptive noise suppression. The system structure and the learning algorithm are developed. The theoretical analysis proves the convergence properties in certain statistical sense. Finally, the numerical results of our system are presented for both standard and the new shrinkage function and compared with the conventional linear adaptive filter based techniques. Results indicate

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

that both optimal solution and the learning performance are superior to the conventional methods. It is also shown that our new shrinkage function performs better than the standard shrinkage function.

2. ADAPTIVE NONLINEAR NOISE SUPPRESSION BASED ON WAVELET SHRINKAGE

2.1 A New Class Of Differentiable Shrinkage Functions

The conventional soft-thresholding function is a continuous function with discontinuous derivative. For optimization problems, the continuous derivative or higher order derivative is often desired. Furthermore, the discontinuous derivative does not allow for optimal solution using other higher order measures, such as Soblov norm, as risk function. Motivated by the differentiable *sigmoid* function which replaces the undifferentiable hard-limited function in traditional neural network, a new class of nonlinear shrinkage functions which have continuous derivatives are constructed as follows:

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

Obviously, $\eta_k(x,t)$ has continuous derivative. Note that when $k \to \infty$, $\eta_k(x,t)$ is just standard soft thresholding function $\eta_s(x,t)$. Those shrinkage functions $\eta_k(x,t)$ will have better numerical properties as will be shown in the examples.

2.2 Adaptive Noise Suppression Using Wavelet Shrinkage

The noise suppression problem in signal processing is similar to but essentially different with function estimation. As known, it can be formulated as follows. Assume a random signal s_i is transmitted over a channel to a sensor that receives the signal with an additive uncorrelated noise n_i. Then the received signal y_i is given by $y_i = s_i + n_i,\ i \in \boldsymbol{Z}$, where Z denotes the integer set. In the wavelet domain (after the DWT), the signal coefficients series is denoted as v_i, the noise coefficients series is denoted as z_i and the received signal coefficients series is denoted as u_i, where $u_i = v_i + z_i,\ i \in \boldsymbol{Z}$. Note that the wavelet coefficients can be easily calculated in real-time using pyramid filter banks [7, 8]. Assume $v(j,k)$ represents the k-th wavelet coefficient at scale j, $j = 1,...,J$, and $v(0,k)$ represents the k-th scaling coefficients at scale J. Time series v_i can be constructed as

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

Similarly, other wavelet coefficients time series can also be constructed. Assume $\hat{s}_i$ is the output estimation of s_i, and $\hat{s}_i$, s_i, n_i are statistically stationary. For orthogonal DWT, the risk function is

$$J(\boldsymbol{t}) = \frac{1}{2} E\{(\hat{s}_i - s_i)^2\} = \frac{1}{2(J+1)} \sum_{j=0}^{J} E\left\{\left(\hat{v}(j,k) - v(j,k)\right)^2\right\}, \tag{2}$$

where the estimated wavelet coefficients $\hat{v}$ are obtained using shrinkage of wavelet coefficients u of received signal, i.e., $\hat{v}(j,k) = \eta_k(u(j,k), t_j)$.

In conventional linear adaptive filtering scheme, the filter coefficients are selected adaptively toward the optimal solution. Similarly, in our nonlinear adaptive noise suppression, we will attempt to select the parameter $\boldsymbol{t}$ in the nonlinear thresholding function $\eta_k(x,t)$ adaptively toward the optimal solution. Note that the scaling coefficients $u(0,k)$ are normally kept without shrinkage, and $\boldsymbol{t}$ denotes vector $[t_1, t_2, \cdots, t_J]^T$ in scale dependent thresholding scheme.

We develop a neural network structure to implement the adaptive noise suppression in wavelet domain using wavelet shrinkage scheme. The adaptive system structure of the proposed method is shown in Fig. 1. In Fig. 1, y' is the reference signal, and u' is the reference wavelet coefficients series in wavelet domain. The estimation $\hat{v}_i$ of v_i is calculated using nonlinear shrinkage functions. The error series $\varepsilon_i = \hat{v}_i - u_i$ in wavelet domain. The following adaptive learning algorithm for neural network so that the risk $J(\boldsymbol{t})$ in (2) can be minimized by adjusting parameter $\boldsymbol{t}$ in real-time.

Step 1. Initialize parameter $\boldsymbol{t} = \boldsymbol{t}_0$.

Step 2. Each time a new sample in wavelet domain u_i and $\hat{v}_i$ is encountered, adjust $\boldsymbol{t}$ using following scheme,

$$\boldsymbol{t}(i+1) = \boldsymbol{t}(i) - \Delta\boldsymbol{t}(i), \tag{3}$$

where

$$\Delta\boldsymbol{t}(i) = \boldsymbol{\alpha}(i) \cdot \frac{\partial \hat{v}_i}{\partial \boldsymbol{t}} \cdot \varepsilon_i, \tag{4}$$

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

In this way, the risk $J(\boldsymbol{t})$ can be minimized and the optimal parameter $\boldsymbol{t}^*$ can be obtained adaptively. However, in practice, the signal time series s_i is unknown. Using s_i as reference signal series y'_i is impractical. In our scheme, we produce the reference signal y'_i with the same transmitted signal time series s_i plus noise series n'_i, which is statistically uncorrelated with noise series n_i, i.e., $y'_i = s_i + n'_i$. This is easy to implement by receiving the signal in another channel.

3. THE CONVERGENCE PERFORMANCE OF THE LEARNING ALGORITHM

Similar to the traditional linear adaptive filtering techniques [1, 2], the analysis of the algorithm will be based on stationary signals, although our nonlinear adaptive filtering methods are designed to track nonstationary random input. This is a usual idealization used so that the analysis becomes relatively tractable. In following theorems, the convergence properties of the adaptive algorithm described in (3) and (4) are given.

Theorem 1 *Suppose a random signal time series* s_i *and a noise time series* n_i *are statistically stationary and noise* n *is a white Gaussian random process with distribution* $N(0,\sigma)$. *If* $\eta_s(x,t)$ *is used as nonlinear shrinkage function, then there exists at most one optimal solution* $\boldsymbol{t}^*$ *which minimize* $J(\boldsymbol{t})$ *in (2) and its every component* $t_j^* \geq 0$.

Theorem 2 *Suppose there exists optimal* $\boldsymbol{t}^*$ *for* $J(\boldsymbol{t})$. *Assume reference signal is* $y'_i = s_i + n'_i$, *where* n' *is a white normal random process which is statistically uncorrelated with* n. *Using the wavelet shrinkage based learning algorithm described in the above section (see (3) and (4)) and* $\eta_s(x,t)$ *as nonlinear shrinkage function, then* $\lim_{i\to\infty} E\{\boldsymbol{t}(i)\} = \boldsymbol{t}^*$, *when* $\alpha(i)$ *is suitable selected, i.e., the learning algorithm is convergent in the mean.*

The proofs of the above two theorems are in [9].

4. NUMERICAL EXAMPLES

The test signals: Blocks, Bumps, HeaviSine and Doppler, which were used by most of other wavelet shrinkage related literature, are shown in Fig. 2. The signal length is 2048 samples. These series are used for adaptive signal processing, i.e., only the past samples before the current time of received signals are assumed to be known. Assume the signal to noise ratio, $SNR = \sigma_s / \sigma_n = 7$, for all signals and the reference signals $y'_i = s_i + n'_i$. Duabechies 16-tap wavelet filter is used. For traditional adaptive noise canceling (ANC) scheme, we take the commonly used reference noise signal $r = n + n'$ [2]. (Using Weiner filter theory, it is easy to find that the ideal performance will be worse if y' is used as a reference signal for ANC.) The 16-tap linear filter is chosen for simulation.

In Table 1, the mean square errors (MSE) of different adaptive methods are given. WANS0 and WANS1 represents the estimated MSE when i = 2048 by wavelet shrinkage based adaptive noise suppression (WANS), using standard soft-thresholding function $\eta_s(x,t)$ and the proposed nonlinear function $\eta_3(x,t)$, respectively. ANC represents the estimated MSE when i = 2048. SOPT0 and SOPT1 represent the optimal MSE by numerical method for scale dependent threshold selection, using $\eta_s(x,t)$ and $\eta_3(x,t)$, respectively. ANC-OPT represents optimal MSE by the ideal linear Weiner filter using ANC scheme. In Fig. 3, the dotted line, dashed line and solid line represent the learning curves of ANC, WANS0 and WANS1 schemes, respectively.

It is indicated in Table 1 that the optimal solutions of the new system (SOPT0, SOPT1) are much better than ANC (ANC-OPT). The learning results of the new system (WANS0, WANS1) are very close to the optimal solutions while the learning results of ANC are not as close to the optimal solutions. In Fig. 3, it is also shown that practical learning performances of the proposed methods are much better than conventional ANC scheme (based on linear adaptive filtering technique). It is also clearly shown that our new shrinkage function (SOPT1 and WANS1) outperforms the standard shrinkage function in both optimal solutions and learning performance. Further numerical results show the similar performance.

5. CONCLUSION

In this paper, a new type of nonlinear adaptive system for adaptive noise suppression based on wavelet shrinkage scheme and a new class of nonlinear shrinkage functions were presented. A new type of wavelet shrinkage based neural network and its adaptive learning algorithm are developed to construct the adaptive system for noise suppression. The theoretical analysis results of convergence properties for the learning algorithm are presented. The numerical results show that the proposed methods perform much better than conventional linear filter based method for a wide range of signals. The further development based on the similar idea but different risk functions, such as *SURE risk*, etc., have also been investigated in [9]. Both theoretical analysis and numerical results show that this new type of adaptive system is very effective for adaptive processing related applications.

6. REFERENCES

[1] S. Haykin, *Adaptive Filter Theory*, Prentice-Hall Information and System Science Series. Prentice-Hall, Englewood Cliffs, New Jersey, 1986.

[2] B. Widrow and S. D. Stearns, *Adaptive Signal Processing*, Prentice-Hall Signal Processing Series. Prentice-Hall, Englewood Cliffs, New Jersey, 1985.

[3] S. Hosur and A. H. Tewfik, “Wavelet transform domain adaptive fir filtering,” *IEEE Trans. on Signal Processing*, vol. 45, no. 3, pp. 617--630, Mar. 1997.

[4] N. Erdol and F. Basbug, “Wavelet transform based adaptive filters: analysis and new results,” *IEEE Trans. on Signal Processing*, vol. 44, no. 9, pp. 2163--2171, Sept. 1997.

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

[6] 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.

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

[8] G. Strang and T. Nguyen, *Wavelets and Filter Banks*, Wellesley-Cambridge Press, Wellesley, 1996.

[9] X.-P. Zhang and M. Desai, “Wavelet shrinkage based adaptive noise suppression using neural network,” *to be submitted*.

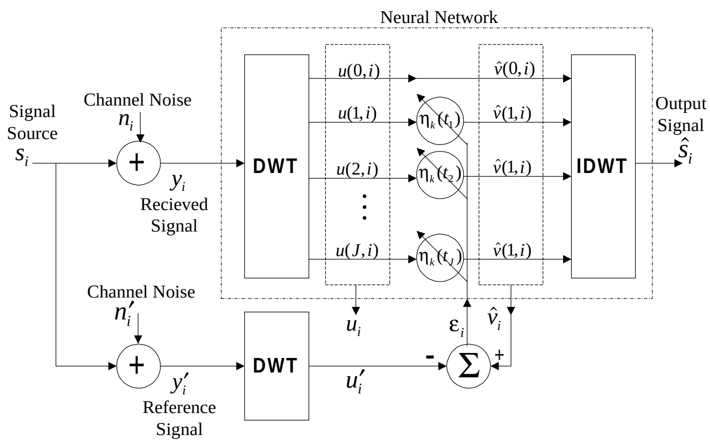

Fig. 1 The adaptive system for noise suppression

	Blocks	Bumps	HeaviSine	Doppler
Original	1.0000	1.0000	1.0000	1.0000
WANS0	0.2548	0.2877	0.1309	0.1666
WANS1	0.2429	0.2812	0.1063	0.1427
ANC	0.5535	0.7382	0.7883	0.7313
SOPT0	0.2306	0.2768	0.0690	0.1158
SOPT1	0.2305	0.2763	0.0616	0.1157
ANC-OPT	0.5	0.5	0.5	0.5

Table 1: MSEs of difference adaptive method

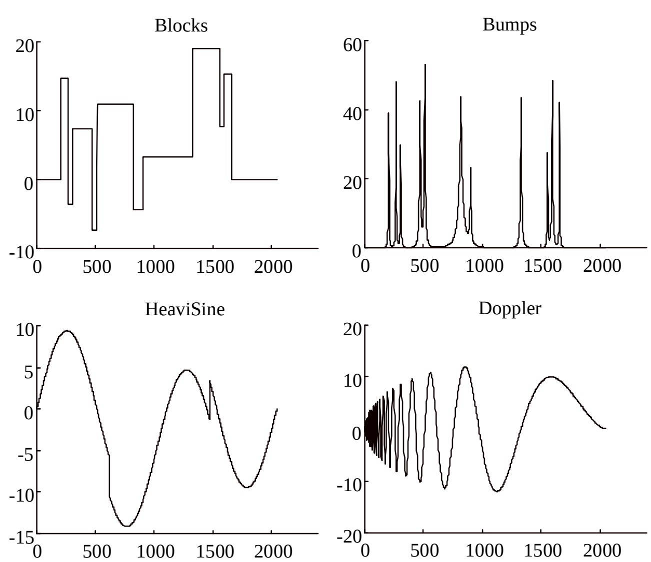

Fig. 2. Four test signals

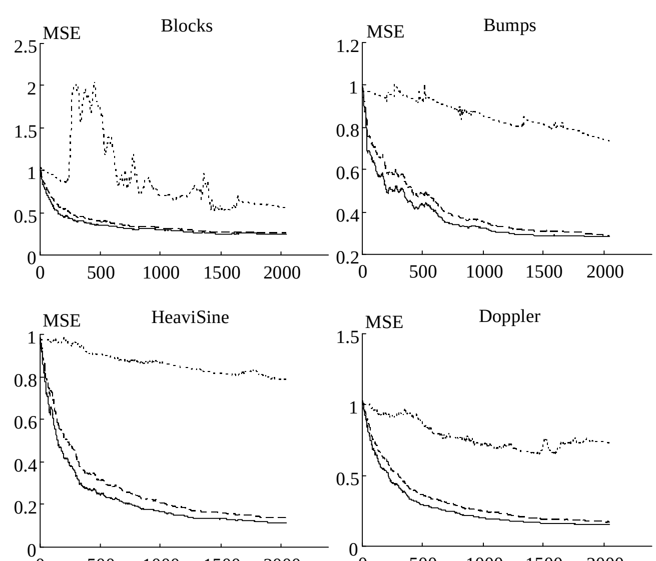

Fig. 3. Learning curves of different adaptive methods