Adaptive Sparse System Identification Using Wavelets

K. C. Ho, *Senior Member, IEEE,* and Shannon D. Blunt, *Member, IEEE*

***Abstract*—This paper proposes the use of wavelets for the identification of an unknown sparse system whose impulse response (IR) is rich is spectral content. The superior time localization property of wavelets allows for the identification and subsequent adaptation of only the nonzero IR regions, resulting in lower complexity and faster convergence speed. An added advantage of using wavelets is their ability to partially decorrelate the input, thereby further increasing convergence speed. Good time localization of nonzero IR regions requires high temporal resolution while good decorrelation of the input requires high spectral resolution. To this end we also propose the use of biorthogonal wavelets which fulfill both of these two requirements to provide additional gain in performance.**

The paper begins with the development of the wavelet-basis (WB) algorithm for sparse system identification. The WB algorithm uses the wavelet decomposition at a single scale to identify the nonzero IR regions and subsequently determines the wavelet coefficients of the unknown sparse system at other scale levels that require adaptation as well. A special implementation of the WB algorithm, the successive-selection wavelet-basis (SSWB), is then introduced to further improve performance when certain *a priori* knowledge of the sparse IR is available. The superior performance of the proposed methods is corroborated through simulations.

***Index Terms*—Adaptive algorithm, sparse impulse response (IR), system identification, wavelets.**

I. Introduction

ADAPTIVE filtering algorithms are finding much widespread use nowadays when the exact nature of a system is unknown or its characteristics are time-varying [1], [2]. Some applications of adaptive filters, such as network echo cancellation, have an impulse response (IR) that contains a large number of zero coefficients. This has led to the development of sparse adaptive algorithms that attempt to minimize the number of adapting coefficients in order to increase the convergence speed, decrease the excess mean-square error (EMSE) and reduce the computational complexity.

Many previous sparse algorithms employed a time domain approach to exploit the sparse nature of a system. Kawamura and Hatori [3] developed a method of sequentially adapting different sets of coefficients to ascertain which ones possess significant nonzero values. While computationally simple, this method may suffer a little in convergence speed due to the sequential adaptation of coefficients. Etter *et al.* [4]–[8] developed an algorithm by which small length filters are placed around delay regions that are estimated to contain significant nonzero coefficients. This algorithm, though faster, may be limited by the increase in complexity caused by adaptive determination of the delay regions in addition to filter coefficients. More recently, Homer *et al.* [9], [10] has formulated an algorithm that is based on statistical detection criterion to detect nonzero IR tap coefficients in the time domain, where the detection threshold is proportional to the power in the desired response. Duttweiler [11] has also recently proposed a technique called the proportionate normalized least mean square (PNLMS) algorithm in which the effective step size for a given tap is proportional to its magnitude to speed up convergence. While this approach exhibits faster convergence than NLMS when the IR is sparse, its performance degrades for nonsparse IR. Different from most sparse algorithms, PNLMS updates all the filter coefficients at each iteration and, therefore, has complexity the same as NLMS. In [12], Benesty and Gay propose a modification to PNLMS that alleviates its susceptibility to nonsparse IR.

This paper proposes sparse system identification in the wavelet domain. Wavelets have very good time localization property [13] and, hence, can locate the nonzero portions of a sparse IR accurately for adaptation. This is accomplished by exploiting the hierarchical structure of wavelets such that, by determining the wavelets at one level of scale whose time domain spans overlap the nonzero IR regions, we can ascertain the wavelets at all other levels of scale whose time domain spans will overlap these nonzero regions as well. As a result, only the unknown system coefficients corresponding to those wavelets are needed for adaptation. Wavelets also have the additional benefit of good spectral localization that provides a certain amount of decorrelation for colored input. The combination of these leads to a dramatic convergence speed improvement while maintaining a computational complexity below that of LMS for sparse system identification.

The idea of using wavelets in adaptive system identification is not new. Much of the previous work, however, considered nonsparse systems and focused on the frequency band decomposition property of wavelets to speed up convergence for correlated input and the fast wavelet transform algorithm to reduce complexity, for example [14]. Another previous approach employs wavelets to reduce the number of coefficients needed to model an IR [15]. This technique requires *a priori* knowledge about the location, shape and duration of the IR. None of this previous work explores the time localization property of wavelets.

Manuscript received May 25, 2001; revised August 2, 2002. This paper was recommended by Associate Editor L. Yong.

K. C. Ho is with the Department of Electrical and Computer Engineering, University of Missouri-Columbia, Columbia, MO 65211 USA (e-mail: hod@missouri.edu).

S. D. Blunt was with the Department of Electrical Engineering, University of Missouri-Columbia, Columbia, MO 65211 USA. He is now with the Radar Division, U.S. Naval Research Laboratory, Washington DC 20375 USA (e-mail: blunt@radar.nrl.navy.mil).

Digital Object Identifier 10.1109/TCSII.2002.807263

For sparse system applications the utilization of this property can lead to a substantial performance gain.

The proposed wavelet-basis (WB) algorithm is an extension of a previously investigated technique called the Haar-Basis (HB) algorithm [16] that has been shown to provide a dramatic convergence speed improvement over the LMS and the Haar transform domain LMS. The HB algorithm utilizes the Haar transform for the identification of nonzero IR regions for adaptation. The HB algorithm can be considered a special case of the WB algorithm in which the Haar wavelet is used. The Haar wavelet has very fine time localization ability but poor spectral localization. The WB algorithm developed here allows the use of any wavelet, including biorthogonal, so that we can gain greater spectral localization and, therefore, greater decorrelation of colored input to speed up convergence. The price paid for greater decorrelation is a slight increase in the number of adapting coefficients because of the decrease in temporal resolution. However, it will be shown later that this tradeoff leads to faster convergence for highly correlated input.

This paper also introduces a new implementation stemming from the structure of the WB algorithm. We shall call it the successive-selection WB (SSWB) algorithm and it will converge even faster than the WB algorithm when the nonzero regions of the sparse IR are impulse-like and widely separated. This kind of IR occurs quite often in many applications, such as those involving multipath and delay channels. The increase in convergence speed arises from a process of selecting nonzero coefficients at successively finer levels of temporal resolution. This allows some of the zero coefficients that would have been selected for adaptation by the WB algorithm to be ignored and, therefore, reduces the total number of adapting coefficients that leads to faster convergence.

One limitation of the proposed algorithms is that they are designed for a sparse IR that is rich in frequency content, i.e., nearly flat frequency response. Rich in frequency content here, generally speaking, refers to an impulse response where over a frequency range from $\pi/8$ to $7\pi/8$, the correlation coefficient of its magnitude response with a constant (flat magnitude spectrum) is bigger than 0.8. This is because, a necessary and not sufficient condition for the proposed method to identify nonzero impulse response region is that, a certain amount of the IR spectral energy must be present in the passband defined by the wavelets. This class of IR is quite common in practice, such as in network echo cancellation application. Annex D of ITU recommendation G.168: Digital Network Echo Cancelers [17] contains several typical echo-path IRs and they are rich in spectral content.

This paper is organized as follows. Section II develops the WB algorithm for sparse system identification. Different kinds of wavelets, including orthogonal and biorthogonal and their relative advantages for sparse system identification are discussed. Section III presents the SSWB algorithm and analyzes its performance. It also summarizes alternative wavelet decompositions that may yield better performance. Section IV provides simulation results to support the theoretical results and illustrate the performance advantages of the proposed methods. Finally, Section V is the conclusion.

II. Sparse Algorithms Using Wavelets

Wavelets possess the attractive property of localization in both the temporal and spectral domains. It is well known that the spectral localization property allows greater decorrelation of colored input and, therefore, enables faster convergence for adaptive algorithms. However, the potential benefits afforded by the time localization property have not previously been exploited.

We shall introduce in this paper the use of the time localization property of wavelets for sparse system identification. The wavelet decomposition at a certain scale $\boldsymbol{a}$ can identify the nonzero regions of a sparse IR at a temporal resolution roughly equal to the duration of the wavelet at scale $\boldsymbol{a}$. In addition, the regular hierarchical structure of the wavelet decomposition permits easy identification of the wavelets at all other levels of scale whose temporal spans overlap the nonzero regions as well. In algorithmic form we can interpret this as, adapting the complete set of filter coefficients corresponding to the wavelet decomposition of the sparse system at a certain scale $\boldsymbol{a}$, determining which of these filter coefficients significantly differ from zero and then activating for adaptation the filter coefficients at all other levels of scale whose corresponding wavelet spans fall into the same IR regions as the detected nonzero coefficients at scale $\boldsymbol{a}$. The filter coefficients associated with wavelets at scale $\boldsymbol{a}$ are called the "control" coefficients. The coefficients corresponding to higher and lower levels of scale are denoted as "parent" and "children" coefficients, respectively.

In order to properly formulate the sparse algorithms to be introduced in the next section, we now present some pertinent details of wavelets for the proposed algorithms.

A. Wavelets

The wavelet decomposition of a signal or system into dyadic scales is accomplished with the use of a low-pass filter and a bandpass filter, denoted as $\boldsymbol{g}$ and $\boldsymbol{h}$, respectively [13]. By repeating a process of convolution and down-sampling of the filter output, we can decompose a signal into many frequency bands at different levels of temporal/spectral resolution. An important factor governing the localization ability of temporal/spectral decomposition is the length of the two filters. Longer wavelet filters will produce greater spectral resolution and lower temporal resolution. More specifically, a longer wavelet filter results in better frequency band sharpness but less knowledge about the exact location of the nonzero IR regions.

In practice, filtering and down-sampling will generate the wavelet decomposition using the least amount of computations and this is often called the fast wavelet transform. For ease of description and analysis purposes we shall use the matrix representation of the wavelet decomposition [13]. It should be noted, however, that the proposed algorithms would, in practice, be implemented using the filtering approach.

We begin with an examination of the system output estimate

$$\hat{y}(k) = \mathbf{w}^T(k)\mathbf{x}(k) \tag{1}$$

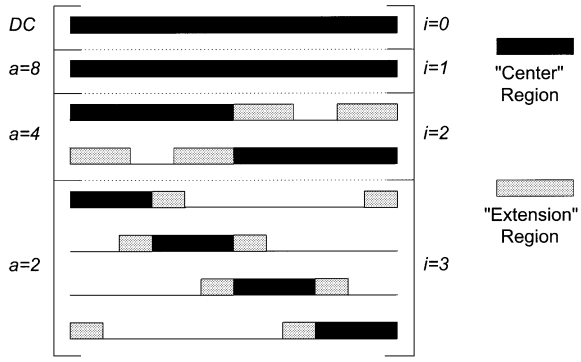

Fig. 1. Wavelet matrix with "centered" wavelets, $N = 8$ and $M = 3$.

where $\mathbf{w}(k)$ is the vector of adaptive coefficients and $\mathbf{x}(k)$ is the vector of input samples, both of which have length N. Without altering the output, $\hat{y}(k)$, we can rewrite (1) as

$$\hat{y}(k) = \mathbf{w}^T(k)\mathbf{Q}^T\mathbf{R}\mathbf{x}(k) \tag{2}$$

where $\mathbf{Q}$ and $\mathbf{R}$ are wavelet matrices. Each row of these matrices, except the first row that represents the dc, contains one wavelet at a particular scale and translation. These matrices result in identity when they are multiplied together as $\mathbf{Q}^T\mathbf{R}$. Then (2) becomes

$$\hat{y}(k) = \mathbf{b}^T(k)\mathbf{z}(k) \tag{3}$$

where $\mathbf{b}(k) = \mathbf{Q}\mathbf{w}(k)$ is interpreted as the vector of wavelet decomposed filter coefficients and $\mathbf{z}(k) = \mathbf{R}\mathbf{x}(k)$ is interpreted as the vector of decomposed input samples.

For orthogonal wavelets, we have $\mathbf{Q} = \mathbf{R}$ and the lowpass and bandpass filters used to generate the wavelet matrix follow an explicit set of rules such that $\boldsymbol{h}$ is the "highpass version" of $\boldsymbol{g}$ [13]. The convergence speed improvement is governed by the tradeoff between decorrelation ability and the number of adapting coefficients, both of which are a function of the lengths of the wavelet filters.

Apart from orthogonal wavelets, we can also use biorthogonal wavelets in which there is no requirement of equality between the wavelet matrices $\mathbf{Q}$ and $\mathbf{R}$ and the strict relationship between $\boldsymbol{g}$ and $\boldsymbol{h}$ has been relaxed [13]. This allows more degrees of freedom to choose the lowpass and bandpass wavelets and, therefore, can lead to greater performance gains in the algorithm. By proper selection of the biorthogonal wavelets we can generate a wavelet matrix $\mathbf{Q}$ in which the lengths of the wavelets at each level of scale are minimized so that the number of adapting coefficients is minimized. At the same time, we can generate a wavelet matrix $\mathbf{R}$ in which the lengths of the wavelets at each level of scale are maximized in order to improve input decorrelation. Given certain constraints the biorthogonal wavelets must satisfy [13], we now have a greater ability to decorrelate the input while maintaining as few adapting coefficients as possible.

The activation of parent and children coefficients for adaptation can be greatly simplified when arranging the wavelets in $\mathbf{Q}$ and $\mathbf{R}$ properly. To this end, the wavelets are ordered from largest scales to smallest as shown in Fig. 1. Furthermore, the wavelets in each row of the wavelet matrix are viewed as having a "center" region and two "extension" regions. The center regions at a certain scale $\boldsymbol{a}$ do not overlap. The extension regions are the remaining length of the wavelet distributed equally on both sides of the center region as shown in Fig. 1. When arranging the wavelets in this way, the parent and children coefficient selection process can be performed in an elegant, straightforward manner.

Finally, we introduce some notation regarding wavelets that we will use in the algorithm formulation. In order to simplify much of the analysis, we associate with each dyadic level of scale $\boldsymbol{a}$ a scale index i such that $\boldsymbol{a} = 2^{(M-i+1)}$ for $1 \leq i \leq M$, where $2^M = N$ is the total filter length. The dc, or coarsest, level in the wavelet transform is associated with scale index $i = 0$. This is also illustrated in Fig. 1 for three levels of scale.

B. WB Algorithm

We assign the $N = 2^M$ wavelet decomposed filter coefficients of $\mathbf{b}(k)$ the coefficient index numbers from 0 to $(2^M - 1)$, with each coefficient corresponding to a single wavelet in $\mathbf{Q}$. The filter coefficients are grouped into sets according to the level of scale of their corresponding wavelets. Each scale index i, for $1 \leq i \leq M$, has the filter coefficient indices $[2^{i-1} : 2^i - 1]$. The filter coefficient index for scale index $i = 0$ is zero. We shall denote the scale index for the control coefficients as δ, where $2 \leq \delta \leq M$. The scale indices $i < \delta$ are, therefore, associated with parent coefficients while scale indices $i > \delta$ are associated with children coefficients. The set of control filter coefficient indices is represented by Δ. The index set of all parent coefficients activated for adaptation (active parent coefficients) is represented by P. Similarly, the index set of the active children coefficients is represented by C.

The WB algorithm operates as follows:

```
REPEAT
  Adapt control coefficients
  Adapt active parent and children coefficients
  IF (end of adaptation interval reached)
     Compare control coefficients with a
     Threshold
     Activate appropriate parent & children coef-
     ficients
  END IF
```

The structure of the proposed algorithm is depicted in Fig. 2. The vector $\mathbf{x}(k) = [x(k)\ x(k-1) \ldots x(k-N+1)]^T$ is a collection of input samples. $\mathbf{z}_\Delta(k)$, $\mathbf{z}_C(k)$ and $\mathbf{z}_P(k)$ are the wavelet decomposed input sample vectors for the control coefficient vector $\mathbf{b}_\Delta(k)$, the active children coefficient vector $\mathbf{b}_C(k)$ and the active parent coefficient vector $\mathbf{b}_P(k)$, respectively. $\hat{y}(k)$ is the estimated system output, $y(k)$ is the actual system response and $e(k)$ is the error. We shall describe below the individual components of the algorithm.

1) Coefficient Adaptation: The system output estimate, as shown in Fig. 2, can be expressed as

$$\hat{y}(k) = \mathbf{b}_\Delta^T(k)\mathbf{z}_\Delta(k) + \mathbf{b}_C^T(k)\mathbf{z}_C(k) + \mathbf{b}_P^T(k)\mathbf{z}_P(k). \tag{4}$$

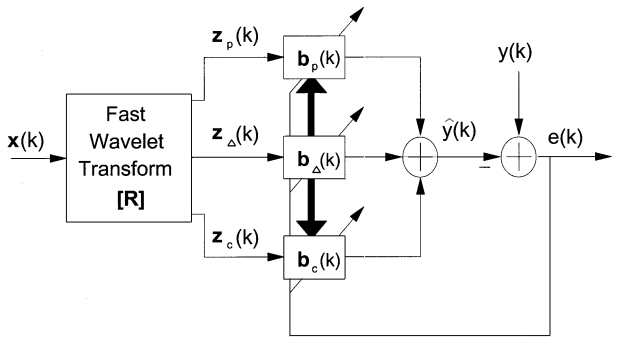

Fig. 2. Block diagram of WB sparse algorithm.

Note that C and P consist of only those children and parent coefficient indices that have been activated for adaptation.

The update of the control coefficients is

$$\mathbf{b}_\Delta(k+1) = \mathbf{b}_\Delta(k) + 2\hat{\mu}e(k)\hat{\lambda}_\delta^{-2}\mathbf{z}_\Delta(k) \tag{5}$$

where $\hat{\lambda}_\delta^{-2}$ is the inverse of the input power estimate for scale index δ, $\hat{\mu} = \mu_0/\hat{N}$, μ_0 is a preset constant and $\hat{N}$ is the current total number of adapting coefficients. Note that all the control coefficients are always adapted. This is to take into account possible changes in the IR so that re-selection of active parent and children coefficients can take place.

For the children and parent coefficients, the update equations are

$$\mathbf{b}_C(k+1) = \mathbf{b}_C(k) + 2\hat{\mu}e(k)\hat{\mathbf{\Lambda}}_C^{-2}\mathbf{z}_C(k) \tag{6}$$

$$\mathbf{b}_P(k+1) = \mathbf{b}_P(k) + 2\hat{\mu}e(k)\hat{\mathbf{\Lambda}}_P^{-2}\mathbf{z}_P(k) \tag{7}$$

where $\hat{\mathbf{\Lambda}}_C^{-2}$ and $\hat{\mathbf{\Lambda}}_P^{-2}$ are diagonal matrices whose diagonal elements are the inverse of the input power estimates at different levels of scale corresponding to the active children and parent coefficients, respectively.

2) Detection of Active Control Coefficients: The coefficients selected for adaptation must be given a suitable period of time to adapt, which we denote as the adaptation interval T_{AI}. T_{AI} can be chosen proportional to the convergence time constant [1]. At the end of each adaptation interval, the algorithm enters into a subroutine that determines which control coefficients are significantly different from zero. Each control coefficient is modeled as a Gaussian random variable [18] with mean equal to the true value and variance governed by the gradient noise. Since the true coefficient value is not known, we use the Neyman-Pearson criteria [19] for detection. For a given probability of false alarm, the detection of nonzero control coefficients is achieved by comparing its magnitude against a threshold. The threshold Φ is formed at the end of each adaptation interval as [16]

$$\Phi = \beta_{fa}\sqrt{\frac{\hat{\mu}\hat{\xi}(k)}{\hat{\lambda}_\delta^2(k)}} \tag{8}$$

where $\hat{\xi}(k)$ is the present mean square error estimate, $\hat{\lambda}_\delta^2(k)$ is the current signal power estimate for scale index δ and they can be obtained by exponential averaging of $e(k)$ and the first element of $\mathbf{z}_\Delta(k)$. β_{fa} is related to the user defined probability of false alarm, P_{fa}, according to the standard unit variance Gaussian and $(\hat{\mu}\hat{\xi}(k)/\hat{\lambda}_\delta^2(k))$ is the estimate of the adaptive coefficient variance [1]. The indices of detected nonzero control coefficients are placed in the coefficient index set Ω.

The probability of detection is dependent on how large the mean of the control coefficients are in comparison with their standard deviations. The standard deviations are proportional to the step size. Decreasing the step size can, therefore, increase the probability of detection, at the cost of a longer convergence time.

3) Parent and Children Coefficient Selection: Once the nonzero control coefficients have been determined, the appropriate parent and children coefficients can be activated. Each detected nonzero control coefficient with index number $\omega \in \Omega$ will activate the children coefficients with indices in the range

$$C_{\omega,i} = \left[2^{(i-\delta)}\omega : 2^{(i-\delta)}(\omega+1) - 1\right] \tag{9}$$

for $(\delta+1) \le i \le M$. This selection process activates children coefficients where the center regions of the corresponding wavelets temporally overlap with the center regions of the wavelets identified by the detected nonzero control coefficients. We consider only the overlap of center regions in order to eliminate unnecessary children activation. This is because all the center regions of the wavelets at a single level of scale together cover the entire time domain impulse response duration. In this manner, we can keep the algorithm simple and also minimize the number of adapting coefficients. The children coefficients activated by all the detected nonzero control coefficients are, as a complete group, designated as $\mathbf{b}_C(k)$.

To find the parent coefficients that need to be activated, we must take the extension regions into account. We first determine those parent coefficients whose corresponding wavelets are so long that they cover most, if not all, of the entire duration of the IR and, therefore, would always be activated. These parent coefficients have scale indices $i = 0, 1, \ldots, r$ where $r = \lceil \log_2(\max(L_g, L_h)) \rceil$ and L_g and L_h are the lengths of the lowpass and bandpass wavelet filters [13]. The total number of parent coefficients in these scales is 2^r and these coefficients may be adapted from the start along with the control coefficients.

Determination of the remaining parent coefficients that must be activated requires two quantities. The first is the duration of the wavelets at scale index i

$$\ell_i = \left(2^{(M-i)} - 1\right) L_g + \left(2^{(M-i)}\right) L_h - \left(2^{(M-i+1)} - 2\right) \quad i > r \tag{10}$$

where $2^M = N$ is the total length of the filter. The second is the temporal translation between two successive wavelets at scale index i

$$k_i = 2^{(M-i+1)}. \tag{11}$$

Then, for each $\omega \in \Omega$, the remaining parent coefficients that need to be activated are those that have index numbers

$$P_\omega = \Big\{ j : [\rho_{\omega,i} - \eta_i] \le j \le [\rho_{\omega,i} + \eta_i], \quad i = (r+1), \ldots, (\delta-1) \Big\} \tag{12}$$

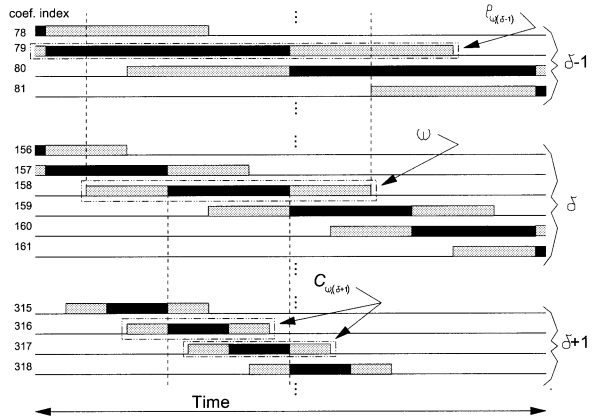

Fig. 3. Parent and children activation for a single control coefficient.

where $\rho_{\omega,i} = \lfloor \omega/2^{(\delta-i)} \rfloor$ and $\eta_i = \lceil (\ell_i - k_i)/(2k_i) \rceil$. $\rho_{\omega,i}$ is the parent coefficient index at scale index i where the center region of the corresponding wavelet overlaps temporally with the nonzero region identified by the detected nonzero control coefficient with index ω. The value η_i is the number of successive center regions at scale index i that overlap with the extension region in one side of a wavelet at scale index i. The parent coefficients activated by all the detected nonzero control coefficients are, as a complete group, designated as $\mathbf{b}_P(k)$.

As an example, Fig. 3 illustrates the case of a single nonzero control coefficient and the necessary coefficients to activate in the nearest (in scale) sets of parents and children. The number of coefficients is $N = 1024$ and the control set scale index is $\delta = 8$. Note that the numbers along the left side of the figure refer to the filter coefficient indices. For the control coefficient $\omega = 158$ detected as nonzero, we activate the children coefficients with indices $C_{\omega,(\delta+1)} = [2(158) : 2(159) - 1] = [316, 317]$ according to (9). Furthermore, for orthogonal wavelets of length $L_g = L_h = 4$, $\rho_{\omega,(\delta-1)} = \lfloor 158/2 \rfloor = 79$ and $\eta_{(\delta-1)} = \lceil (22-8)/((2)(8)) \rceil = 1$ and the parent coefficients with indices [78:80] will be activated.

Finally, note that at scale index i, the parent coefficient indices are in the range $[2^{(i-1)} : 2^i - 1]$. Therefore, any activated parent coefficient indices $[(\rho_{\omega,i} - \eta_i), \ldots, (\rho_{\omega,i} + \eta_i)]$ at scale index i which exceed the highest coefficient index number $(2^i - 1)$, must be decreased by $2^{(i-1)}$. Conversely, any coefficient indices less than the lowest coefficient index number $2^{(i-1)}$, at scale index i, must be increased by $2^{(i-1)}$. This occurs because, in matrix form, the wavelets shift cyclically and some of them cover both ends of the temporal domain. This is illustrated in Fig. 1 for the first and last wavelets at scale index $i = 3$.

C. Expected Number of Adapting Coefficients at Steady State

The excess mean square error (mse) of the proposed algorithm is governed by the number of adapting coefficients at steady state $\hat{N}_{\rm SS}$. $\hat{N}_{\rm SS}$ is a random variable because it is determined by the number of detected nonzero control coefficients which are themselves random variables. We, therefore, evaluate the expected value of $\hat{N}_{\rm SS}$ and it is equal to

$$E[\hat{N}_{\rm SS}] = N_{\rm control} + E\left[\hat{N}_{\rm children,SS}\right] + E\left[\hat{N}_{\rm Parents,SS}\right]. \quad (13)$$

$N_{\rm control} = 2^{(\delta-1)}$ is the number of control coefficients and it is a constant since all of them are always adapted to take care of the possible changes in the unknown IR to be identified. Due to the dyadic structure of the wavelet decomposition, from (9) the number of children coefficients activated by a single nonzero control coefficient is $\sum_{j=1}^{(M-\delta)} 2^j = 2(2^{(M-\delta)} - 1)$, where $\delta < M$. If the number of true nonzero control coefficients is Ψ_δ, then the expected number of adapting children coefficients at steady state is

$$E\left[\hat{N}_{\rm children,SS}\right] = 2\left[2^{(M-\delta)} - 1\right]\Psi_\delta + 2\left[2^{(M-\delta)} - 1\right]\left(2^{(\delta-1)} - \Psi_\delta\right)P_{\rm fa}. \quad (14)$$

The first term is the number of children coefficients activated by the true Ψ_δ nonzero control coefficients and the second term is the expected number of children coefficients activated due to false detection of the remaining $(2^{(\delta-1)} - \Psi_\delta)$ zero control coefficients to be nonzero. Note that parent and children coefficients activated by the false detection of zero control coefficients to be nonzero will stop adaptation and be reset to zero if the false alarm is cleared at the end of the next adaptation interval.

A general formula for the expected number of adapting parent coefficients at steady state is very cumbersome due to its dependence on the temporal distribution of the nonzero IR regions and on the choice of wavelet. However, the number of active parent coefficients is relatively small compared to the number of control and active children coefficients. For simplicity, we approximate the number of active parent coefficients as a function of the number of true nonzero control coefficients. Ignoring the extension regions, a single true nonzero control coefficient will activate one parent coefficient at each scale up to scale level r and, hence, in total $[2^r + (\delta - r)]$ parent coefficients. Extending this to multiple true nonzero control coefficients leads to $\min\left\{2^{(\delta-1)}, [2^r + (\delta - r)\Psi_\delta]\right\}$, where $2^{(\delta-1)}$ is the maximum number of parent coefficients. Finally, upon including the effects of activation by falsely detected nonzero control coefficients, the approximate expected number of parent coefficients is

$$E\left[\hat{N}_{\rm parents,SS}\right] \approx \min\left\{2^{(\delta-1)}, \left[2^r + (\delta - r)\left(\Psi_\delta + \left(2^{(\delta-1)} - \Psi_\delta\right)P_{fa}\right)\right]\right\}$$

After combining like terms, we have according to (13)

$$E\left[\hat{N}_{\rm SS}\right] \approx 2^{(\delta-1)} + 2\left[2^{(M-\delta)} - 1\right] \cdot \left(\Psi_\delta + \left(2^{(\delta-1)} - \Psi_\delta\right)P_{fa}\right) + \min\left\{2^{(\delta-1)}, \left[2^r + (\delta - r) \cdot \left(\Psi_\delta + \left(2^{(\delta-1)} - \Psi_\delta\right)P_{\rm fa}\right)\right]\right\}. \quad (15)$$

In (15) we have ignored the term from the probability of miss. This is because the nonzero control coefficients that are missed will have small magnitudes and in most cases will contribute a negligible residual excess error compared to the noise floor.

For a typical sparse IR occuring in practice, $E[\hat{N}_{\rm SS}]$ will be considerably less than N. For instance, if $N = 1024$ ($M = 10$), $\delta = 7$, $r = 2$ ($L_g = L_h = 4$), $\Psi_\delta = 6$ and $P_{\rm fa} = 0.01$, then $E[\hat{N}_{\rm SS}] \approx 193$ which is five times less than N.

Note that the WB algorithm operates under the assumption that the unknown sparse IR is rich in spectral content. Viewing each level of scale as a filter bank, this ensures that there exists significant IR energy in the portion of the spectrum defined by the control set wavelets. Furthermore, while negligible in practice, it may be possible to contrive situations in which rich spectral content exists yet the IR is orthogonal to the control set wavelets. In such a case this may be alleviated by cycling through a group of different control sets while maintaining adaptation of the selected parent and children coefficients.

III. Special Realizations of the WB Algorithm

The WB algorithm is a general technique for sparse system identification using wavelets. However, there are some special implementations of the WB algorithm that, under certain conditions, yield even greater performance improvement. We examine one such special implementation here. This is achieved by altering the method of children selection to take advantage of high temporal resolution. We denote this as the SSWB algorithm in which the children activation process is performed sequentially. The detected nonzero control coefficients only activate children coefficients in the next children scale index (δ+1). The detected nonzero children coefficients at this scale index are then used to activate the appropriate children coefficients in the next children scale index ($\delta + 2$) and so on. This allows many of the zero children coefficients to be excluded from adaptation. For an IR containing many nonzero regions that are impulse-like and widely distributed over the entire temporal domain, a significant reduction in the number of adapting coefficients can be realized, resulting in a faster convergence speed.

A. SSWB Algorithm

The SSWB algorithm is a special realization of the WB algorithm. The adaptation process and the activation of parent coefficients remains the same. The difference is in how the children coefficients are activated. In the WB algorithm, the children coefficients make up a large portion of the total number of adapting coefficients due to the fact that the number of coefficients doubles as the children scale index increases by one. Therefore, unless the nonzero region of an IR matches up perfectly with the duration of a given control set wavelet, there are unnecessary children coefficients being activated when the nonzero IR regions are impulse-like and widely distributed.

The SSWB algorithm seeks to remedy this by utilizing the detected nonzero filter coefficients at scale index i, where $\delta \leq i \leq (M-1)$, to select the appropriate children coefficients for adaptation at the next scale index (i+1) only. As children coefficients in successively higher scale indices(i.e., higher temporal resolution) are activated, a more accurate picture of the exact location of the nonzero IR regions becomes apparent.

An example is illustrated in Fig. 4 for comparison of the WB and SSWB algorithms where sections of the children coefficient selection are shown. The time domain IR is a single impulse

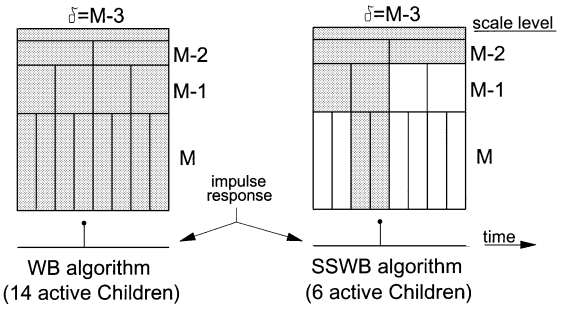

Fig. 4. Children activation for WB and SSWB algorithms.

shown at the bottom of Fig. 4. The control set has scale index $\delta = M - 3$. We can see that while the WB algorithm would require 14 active children coefficients, the SSWB algorithm only requires six active children coefficients. We can also see that if this example is expanded to a more practical situation with a long filter length where multiple impulses may occur, the total number of filter coefficients for the SSWB algorithm can be significantly less than the WB algorithm.

There is one limitation on the SSWB algorithm which relates to the nature of the sparse IR itself. Because the SSWB algorithm successively detects the filter coefficients at finer time resolutions, each nonzero region in the IR must have significant energy in the frequency bands corresponding to the scale indices $\delta \leq i \leq (M-1)$. This is because each nonzero filter coefficient in scale index i must have enough energy to be detectable before it can activate the appropriate children coefficients at the next scale index (i+1). It is, therefore, easy to see that the SSWB algorithm performs best when the IR is impulse-like over short time periods such as in multipath and delay channels.

1) Children Selection for the SSWB Algorithm: The selection of children coefficients for the SSWB algorithm is an extension of the method used for the WB algorithm. Each control coefficient and each active child coefficient is compared with a threshold. This threshold is determined at the end of each adaptation interval before the detection process takes place. A separate threshold is required at each scale index i, for $\delta \leq i \leq (M-1)$ and is chosen as

$$\Phi_i = \beta_{fa}\sqrt{\frac{\hat{\mu}\hat{\xi}(k)}{\hat{\lambda}_i^2(k)}} \tag{16}$$

where $\hat{\lambda}_i^2(k)$ is the current signal power estimate at scale index i and the remaining variables are the same as those previously discussed for the WB algorithm. It is also possible to increase the probability of false alarm and, therefore, decrease $\beta_{\rm fa}$, at higher scale indices to make detection easier with little detriment to the number of adapting coefficients. Each detected nonzero child coefficient with coefficient index j will activate the children coefficients with coefficient indices $2j$ and $2j+1$ only.

2) Expected Number of Adapting Coefficients at Steady State: This subsection determines the expected number of adapting coefficients at steady state for the SSWB algorithm. We need only to examine the children coefficients since the number of control coefficients and method of parent coefficient activation remains the same as the WB algorithm.

We shall denote the number of adapting coefficients at scale index i as $\hat{N}_i$ and the number of true nonzero coefficients at scale index i as Ψ_i. Each true nonzero control coefficient will activate two children coefficients at scale index $i = (\delta + 1)$. Also, due to the finite probability of false detection there may be coefficients whose true value is zero but are falsely detected as nonzero. Combining these two quantities and taking expectation over time gives, at steady state, the expected number of adapting coefficients at scale index $(\delta + 1)$ as $E[\hat{N}_{(\delta+1),\mathrm{SS}}] = 2\Psi_\delta + 2P_{\mathrm{fa}}(E[\hat{N}_\delta] - \Psi_\delta)$, which after rearranging becomes

$$E[\hat{N}_{(\delta+1),\mathrm{SS}}] = (2 - 2P_{\mathrm{fa}})\Psi + 2P_{\mathrm{fa}}E[\hat{N}_\delta] \tag{17}$$

where $E[\hat{N}_\delta] = \hat{N}_\delta = 2^{(\delta-1)}$ since all control coefficients are always adapted. Similarly, at scale index $(\delta + 2)$ we have $E[\hat{N}_{(\delta+2),\mathrm{SS}}] = (2-2P_{\mathrm{fa}})\Psi_{(\delta+1)} + 2P_{\mathrm{fa}}E[\hat{N}_{(\delta+1),\mathrm{SS}}]$. Substituting $E[\hat{N}_{(\delta+1),\mathrm{SS}}]$ in (17) yields

$$E\left[\hat{N}_{(\delta+2),\mathrm{SS}}\right] = (2 - 2P_{\mathrm{fa}})\,\Psi_{(\delta+1)} + (2 - 2P_{\mathrm{fa}})\,(2P_{\mathrm{fa}})\,\Psi_\delta + (2P_{\mathrm{fa}})^2\,E\left[\hat{N}_\delta\right]. \tag{18}$$

Generalizing the result to children scale index i yields

$$E\left[\hat{N}_{\mathrm{i,SS}}\right] = (2 - 2P_{\mathrm{fa}}) \sum_{\ell=0}^{i-\delta-1} (2P_{\mathrm{fa}})^\ell\,\Psi_{(i-\ell-1)} + (2P_{\mathrm{fa}})^{(i-\delta)}\,E\left[\hat{N}_\delta\right]. \tag{19}$$

Combining the result for $(\delta + 1) \le i \le M$ leads to the total expected number of children coefficients at steady state, $E[\hat{N}_{\mathrm{children,SS}}]$, as

$$E\left[\sum_{i=\delta+1}^{M} \hat{N}_{\mathrm{i,SS}}\right] = (2 - 2P_{fa}) \sum_{i=\delta+1}^{M} \sum_{\ell=0}^{i-\delta-1} (2P_{\mathrm{fa}})^\ell\,\Psi_{(i-\ell-1)} + E\left[\hat{N}_\delta\right] \sum_{i=\delta+1}^{M} (2P_{fa})^{(i-\delta)}. \tag{20}$$

This can be greatly simplified if we make a change of variables for the summation indices and apply the valid assumption that P_{fa} is small and, therefore, the higher order terms of P_{fa} are negligible. As a result

$$E\left[\hat{N}_{\mathrm{children,SS}}\right] \approx 2^\delta P_{\mathrm{fa}} + (2 - 2P_{\mathrm{fa}})\,\Psi_{(M-1)} + (2 + 2P_{\mathrm{fa}}) \sum_{j=\delta}^{M-2} \Psi_j \tag{21}$$

which is a more tractable solution. Combining this with the number of control coefficients and the approximate number of parent coefficients from Section II-C gives the total expected number of adapting coefficients for the SSWB algorithm

$$E\left[\hat{N}_{\mathrm{SS}}\right] \approx (1 + 2P_{\mathrm{fa}})\,2^{(\delta-1)} + (2 - 2P_{\mathrm{fa}})\,\Psi_{(M-1)} + (2 + 2P_{\mathrm{fa}}) \sum_{j=\delta}^{M-2} \Psi_j + \min\left\{2^{(\delta-1)}, \left[2^r + (\delta - r) \cdot \left(\Psi_\delta + \left(2^{(\delta-1)} - \Psi_\delta\right) P_{\mathrm{fa}}\right)\right]\right\}. \tag{22}$$

A numerical analysis of (22) reveals a tendency for $E[\hat{N}_{\mathrm{SS}}]$ to significantly decrease as the control set scale index δ decreases (increasing level of scale). It should be noted, however, that while the theoretical number of adapting coefficients can get very small as δ decreases, there is a practical limit on δ at which the algorithm will no longer converge. This is because the spectral bandwidth decreases by a factor of two as the scale index decreases by 1. This will typically result in less IR energy as the scale index decreases which translates into smaller filter coefficient values. Smaller coefficients are difficult to detect and will hinder convergence.

B. Alternative Implementations of the WB Algorithm

The WB and SSWB algorithms employ the dyadic wavelet decomposition to achieve temporal localization of nonzero IR regions as well as a certain amount of input decorrelation. It may be possible to achieve even greater performance improvement by using a wavelet packet approach to tailor the form of the wavelet decomposition to best fit the application of interest. For instance, when the nonzero coefficients of a sparse IR are highly localized in the time domain, high temporal resolution is not necessary. In this case, decomposing the IR so as to improve the spectral resolution can result in improved input decorrelation while maintaining temporal localization.

One such approach is to employ an equal-resolution decomposition which is similar to a short-term Fourier transform except for the temporal localization ability provided by wavelets. In this way, all the sets of wavelets have the same level of scale and better spectral resolution is achieved at the high end of the spectrum. However, care must be taken because the finite wavelet span may cause significant out-of-band sidelobes to occur in the magnitude spectrum of some wavelet at certain scales. This can hinder decorrelation resulting in still-small eigenvalues in the transformed input correlation matrix. The modes of coefficient convergence are inversely proportional to the eigenvalues [1]. Hence, there could be slow convergence in the mean-square difference [1]

$$\mathrm{MSD} = E\left[(\mathbf{b} - \mathbf{b}^*)^T\,(\mathbf{b} - \mathbf{b}^*)\right] \tag{23}$$

where $\mathbf{b}$ is the vector of adapting coefficients and $\mathbf{b}^*$ is its true value. However, some performance improvement is still possible for the EMSE [1]

$$\mathrm{EMSE} = E\left[(\mathbf{b} - \mathbf{b}^*)^T\,\mathbf{R}_z\,(\mathbf{b} - \mathbf{b}^*)\right] \tag{24}$$

where $\mathbf{R}_z$ is the input correlation matrix in the wavelet domain. This is because the large coefficient errors from the slower modes of convergence are weighted by their respective eigenvalues that are themselves very small. Since the eigenvalues are inherent to the autocorrelation matrix, the normalization factors

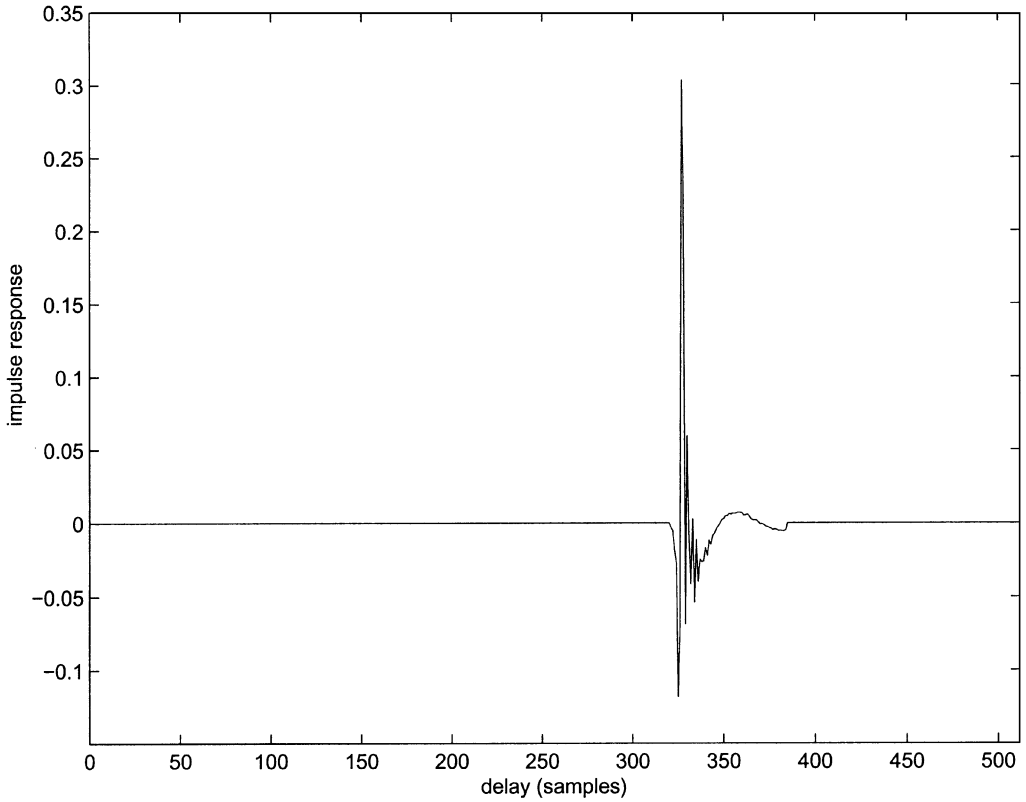

Fig. 5. Echo path IR $M_1(k)$.

are effectively canceled out so that a larger step size may be employed, enabling faster convergence. In other words, this manner of decomposition may work well when the interest is the error signal such as in echo cancellation application instead of system identification where the coefficient values are of interest.

Another approach recently proposed is a symmetric-resolution approach [20] in which the signal component at the lowest level of scale ($\boldsymbol{a} = 2$) is further decomposed such as to be symmetric in spectral resolution with the other higher levels of scale. This manner of decomposition has demonstrated convergence speed improvement when the input signal has correlation peaks in the upper half of the spectrum (i.e., angular frequency from $\pi/2$ to π). The optimal decomposition through wavelet packets for a given input correlation and IR type is a subject of further study.

IV. Simulation Results

This section presents simulations to evaluate the performance of the proposed sparse adaptive algorithms. We examine the use of different wavelets with the WB algorithm to illustrate the benefit of using orthogonal and biorthogonal wavelets for sparse system identification as compared to LMS. We also use a biorthogonal Haar wavelet in the SSWB algorithm to demonstrate the possible performance gain relative to WB and LMS.

The filter length is $N = 512$ (i.e., $M = 9$) and the probability of false alarm is set to $P_{\text{fa}} = 0.01$. The length of the adaptation interval is set to two adaptation time constants [1]. The time constant for the LMS-based updates (5)–(7) is about $1/(4\hat{\mu})$ and, hence, $T_{\text{AI}} = 2(1/(4\hat{\mu})) = \hat{N}/(2\mu_0)$, where $\hat{N}$ is the number of adapting coefficients and μ_0 is a preset constant step size. Note that $\hat{N}$ varies from one adaptation interval to another and is constant within. Also, we will use $\delta = M - 3$ for the control set, which corresponds to level of scale $\boldsymbol{a} = 16$ and there are 32 control coefficients always being adapted. The performance indices are the sum of squared coefficient error (MSD) shown in (23) and the mse. In each set of simulations, the preset step size constant μ_0 for each of the algorithms is adjusted to provide the same level of steady state MSD or EMSE for all, depending on which is being examined.

A. Behavior of the WB Algorithm

The performance of the WB algorithm is presented to reveal the benefit in using orthogonal and biorthogonal wavelets for sparse system identification. The WB algorithm is a general sparse adaptive algorithm and does not make any assumption about the characteristics of an IR other than it being sparse. We use echo cancellation application as an example to evaluate the performance of the WB algorithm. The input is the lowpass correlated sequence

$$x(n) = 1.8\cos(0.1875\pi)x(k-1) - (0.9)^2 x(k-2) + \varepsilon(k) \quad (25)$$

where $\varepsilon(k)$ is a white Gaussian noise sequence and the noise power is $\sigma_n^2 = 10^{-3}$.

Three different wavelets are examined for the WB algorithm: the Haar wavelet (such that the WB reduces to the HB algorithm) with $L_g = L_h = 2$ denoted as Haar(2,2), the Daubechies' wavelet with $L_g = L_h = 4$ denoted as Daub(4,4) and the biorthogonal Haar type wavelet with $L_g = 2$ and

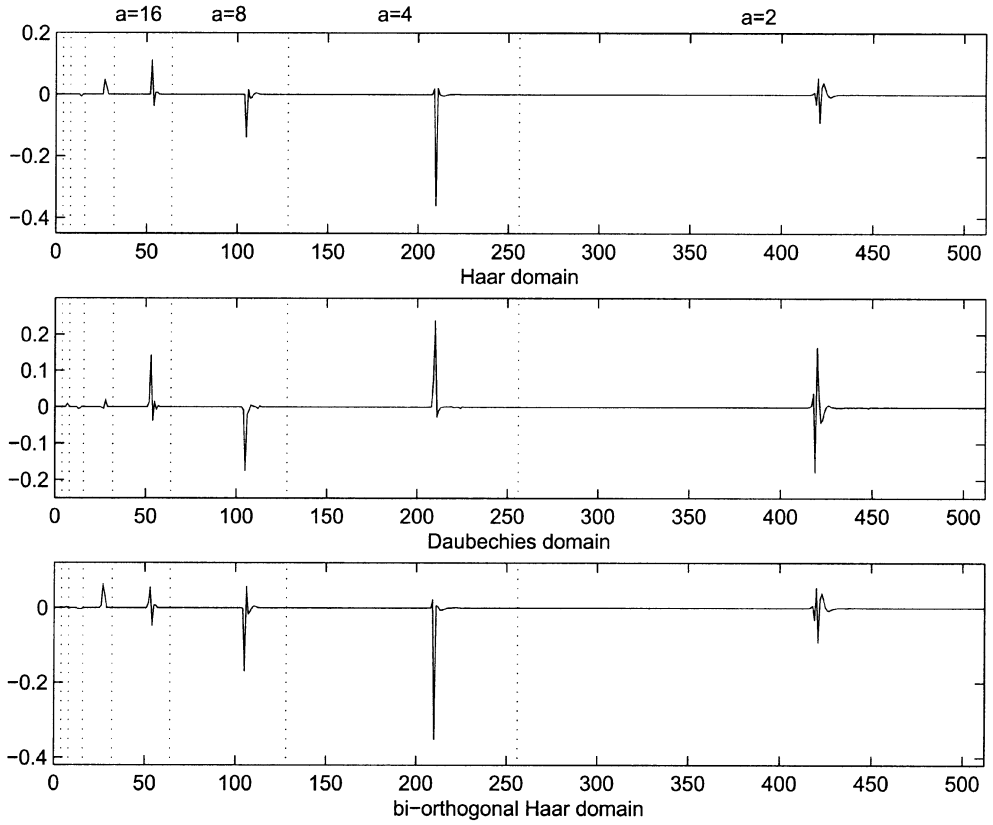

Fig. 6. Echo path IR $M_1(k)$ in the wavelet domains.

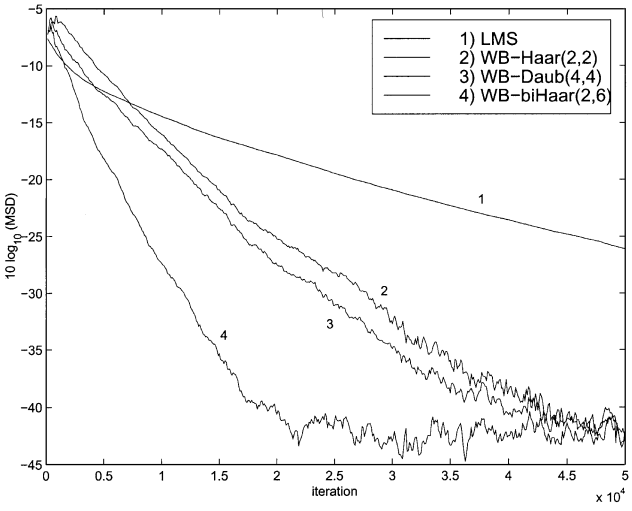

Fig. 7. MSD convergence for WB and LMS.

$L_h = 6$ denoted as biHaar(2,6) [13]. The echo-path IR is $M_1(k)$ from [17] and has 64 nonzero samples with some initial flat delay as illustrated in Fig. 5. The echo-path IR is scaled by a constant to produce an echo return loss (ERL), the ratio of the power between $x(k)$ and $y(k)$, of 6 dB. In the respective wavelet domains, this echo path is represented as in Fig. 6.

The results of the WB algorithm for each of the three wavelets and the LMS algorithm are illustrated in Fig. 7 for the MSD case. The inital slow convergence of the WB algorithm is due to the need to activate the appropriate parent and children coefficients for adaptation. The WB algorithm using Haar(2,2) requires only $E[\hat{N}_{\rm SS}] \cong 101$ coefficients on average at steady state and converges much faster than LMS, which reaches steady state at $\sim 15 \times 10^4$ iterations. Using Daub(4,4) gives better decorrelation of the input and, therefore, converges faster than Haar(2,2) although the average number of adapting coefficients at steady state increases to $E[\hat{N}_{\rm SS}] \cong 140$. BiHaar(2,6) converges significantly faster than Daub(4,4). Bihaar(2,6) applies longer wavelets to the input to give better decorrelation and shorter wavelets to the filter coefficients to minimize the average number of active coefficients at steady state and it gives only $E[\hat{N}_{\rm SS}] = 126$ adapting coefficients in this simulation. $E[\hat{N}_{\rm SS}]$ for each of these choices of wavelets is dramatically less than the 512 required by the LMS algorithm. The approximate expected number of adapting coefficients for Haar(2,2), Daub(4,4) and biHaar(2,6) computed from (15) are 115, 149, and 130, respectively. The estimates are close to the values from simulation. The differences are due to the fact that this IR is very localized and, therefore, the approximate number of parent coefficients will tend to be slightly overestimated.

The results for the mse are illustrated in Fig. 8. We find that the WB using Haar(2,2) again significantly outperforms LMS which requires $\sim 13 \times 10^4$ iterations to reach steady state. Furthermore, the WB with Daub(4,4) is faster than Haar(2,2) and biHaar(2,6) is faster still.

Finally, we provide a simple example to illustrate the tracking ability of the WB algorithm for time-varying systems. To do this we use the biHaar(2,6) wavelet, the correlated input from (25) and the $M_1(k)$ IR. Midway through the simulation

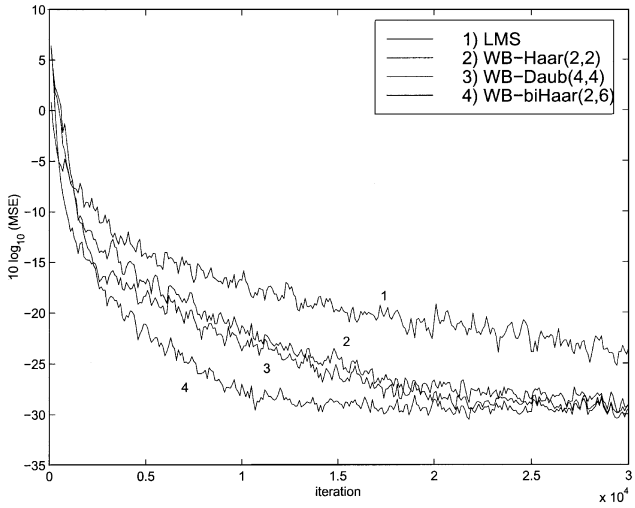

Fig. 8. MSE convergence for WB and LMS.

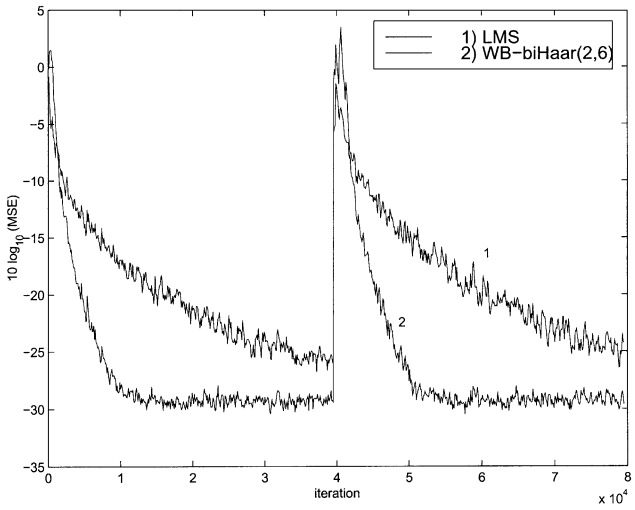

Fig. 9. MSE convergence for WB and LMS under time-varying system.

at 40×10^3 iteration, we change the IR by moving the nonzero portion to a completely different temporal location that does not overlap with its previous location. In this way completely different control coefficients must be detected in order to "locate" the nonzero region. Fig. 9 depicts the performance of the WB and LMS algorithms in this situation and it is obvious that the WB algorithm is able to quickly "relocate" the nonzero IR region and then reconverge much faster than LMS.

B. Behavior of the SSWB Algorithm

The SSWB algorithm uses the same correlated input as generated from (25) and the noise power is set to the same level of $\sigma_n^2 = 10^{-3}$. The IR is chosen as shown in Fig. 10 where the nonzero regions are impulse-like and widely separated. The SSWB algorithm works best for this type of IR. The biHaar(2,6) wavelet is chosen for the SSWB. Fig. 11 shows the performance for the MSD case of the SSWB algorithm relative to LMS and the WB algorithm using biHaar(2,6). The SSWB algorithm converges considerably faster than WB and much faster than LMS which requires $\sim 80 \times 10^4$ iterations to converge. On average, the SSWB algorithm requires only $\hat{N}_{\mathrm{SSWB}} \cong 113$ coefficients at steady state which is less than the $\hat{N}_{\mathrm{WB}} \cong 179$ required for the WB and much less than the 512 required for LMS. The expected number of adapting coefficients for SSWB with biHaar(2,6) calculated from (22) is 113 and for WB with biHaar(2,6) from (15) is 179. Both of these match up well with their actual values from simulations because the IR in this instance will cause the maximum number of parent coefficients to be activated so there is no overestimation. The impulse like IR will make the WB algorithm activate many unnecessary zero children coefficients and, therefore, have slower convergence. SSWB is able to provide much better performance in this case.

The mse convergence curves are illustrated in Fig. 12 for SSWB and WB using biHaar(2,6) and LMS. We notice that the SSWB algorithm converges dramatically faster than WB, which is itself much faster than LMS taking $\sim 40 \times 10^4$ iterations to reach steady state.

C. Computational Complexity

For the WB and SSWB algorithms there will be a slight computational "bottle neck" at the end of each adaptation interval due to the parent and children selection process. Since the adaptation interval is typically long, though, the increase in the average computational load due to the selection process is negligible. Also, this computational increase in most cases can be handled through different implementations in a DSP chip. For example, it can be alleviated by not updating coefficients during the iteration where the selection process takes place.

When neglecting the selection process, we can derive an expression for the complexity during a "typical" iteration. The power estimation through exponential averaging for the update (5), (6), and (7) and the fast wavelet decomposition will require $3\log_2 N$ and $(L_g + L_h)\log_2 N$ operations [13], respectively, where L_g and L_h are the lengths of the lowpass and bandpass wavelet filters. Adding to this the computations required for power normalization and the update of the $\hat{N}$ coefficients, we find that the WB algorithm requires about

$$\begin{aligned} 2\hat{N} + (4 + L_g + L_h)\log_2 N \quad & \text{operations} \\ \log_2 N \quad & \text{divisions} \end{aligned} \tag{26}$$

at each iteration, where one operation is defined as one multiplication and one addition. For a typical sparse IR, the WB algorithm has a lower complexity than LMS. The complexity of the SSWB algorithm is also given by (26). However, due to smaller $\hat{N}$, SSWB will have a lower complexity than WB for any widely distributed, impulse-like IR.

As a simple example, if we use biHaar(2,6) and set $N = 512$ ($M = 9$) and $\delta = M - 3$ and assume that one division requires 10 operations, the WB algorithm will have complexity less than LMS as long as $\hat{N} < 413$. This is often fulfilled for a sparse system.

V. Conclusion

We have proposed an adaptive algorithm for sparse system identification using wavelets for use in systems that are rich in spectral content. The algorithm exploits both the spectral and temporal localization properties of wavelets in order to realize

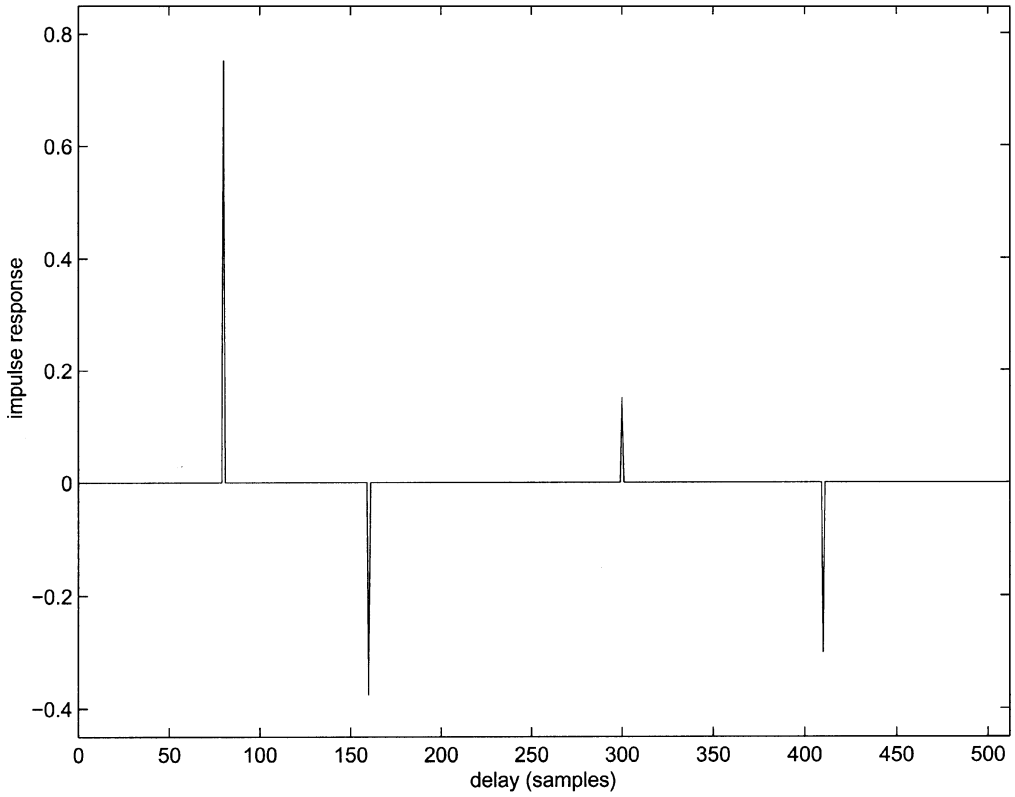

Fig. 10. Dispersed IR with impulse-like nonzero coefficients.

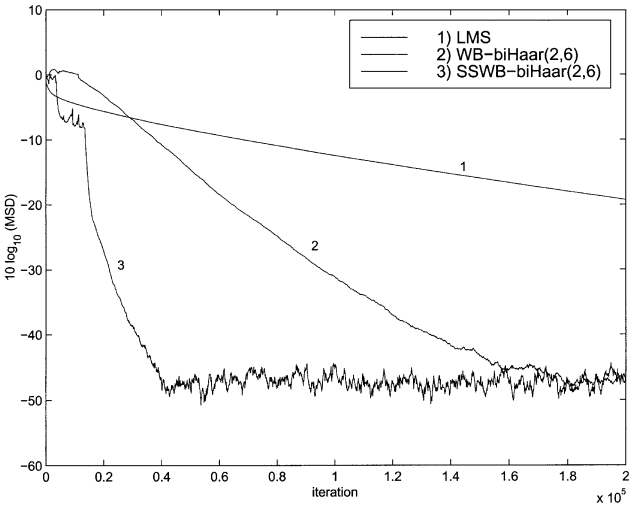

Fig. 11. MSD performance for SSWB, WB, and LMS for impulse like sparse IR.

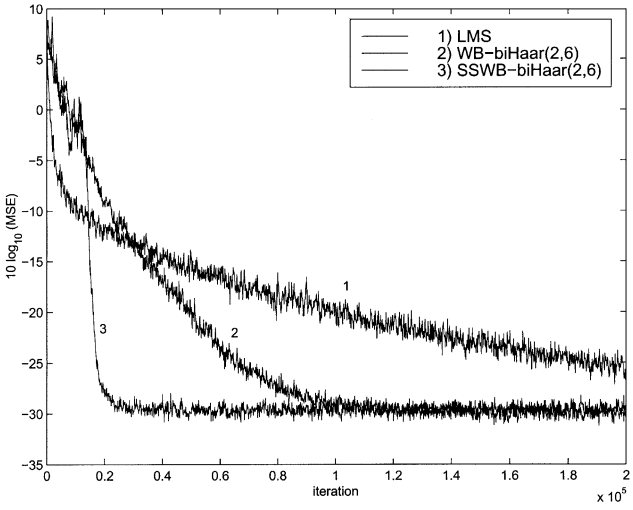

Fig. 12. MSE performance for SSWB, WB, and LMS for impulse like sparse IR.

dramatic performance gains for sparse system identification. The spectral localization property provides partial decorrelation of the input to increase the convergence speed, while the temporal localization property identifies the nonzero IR regions to reduce the number of adapting coefficients and, therefore, reduce complexity and further speed up convergence. This is done by using coefficients corresponding to a single level of scale to identify all remaining coefficients that require adaptation. Furthermore, the proposed WB algorithm supports the use of orthogonal as well as biorthogonal wavelets, which can lead to even greater performance improvement. We have also proposed a special realization of the WB algorithm, the SSWB, that takes advantage of the properties of a certain type of sparse system. The SSWB algorithm produces even more pronounced results than the WB algorithm for sparse systems that have many impulse-like nonzero regions located far apart. The convergence speed improvement for the proposed algorithms is quite dramatic compared to LMS, while at the same time having lower computational complexity than LMS for most practical sparse systems.

ACKNOWLEDGMENT

The authors would like to thank the reviewers for providing valuable comments and suggestions, which have greatly improved the paper.

REFERENCES

[1] S. Haykin, *Adaptive Filter Theory*, 2nd ed. Englewood Cliffs, NJ: Prentice-Hall, 1996.
[2] W. K. Jenkins *et al.*, *Advanced Concepts in Adaptive Signal Processing*, 1st ed. Boston: Kluwer, 1996.
[3] S. Kawamura and M. Hatori, "A tap selection algorithm for adaptive filters," in *Proc. IEEE Int. Conf. Acoustics, Speech Signal Processing*, Atlanta, GA, May 1996, pp. 2979–2982.
[4] P. C. Yip and D. M. Etter, "An adaptive technique for multiple echo cancellation in telephone networks," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, Dallas, TX, Apr. 1987, pp. 2133–2136.
[5] D. M. Etter, "System modeling using an adaptive delay filter," *IEEE Trans. Circuits Syst. II*, vol. CAS-34, pp. 770–774, July 1987.
[6] Y. F. Cheng and D. M. Etter, "Analysis of an adaptive technique for modeling sparse systems," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, pp. 254–264, Feb. 1989.
[7] P. C. Yip and D. M. Etter, "An adaptive multiple echo canceller for slowly time-varying echo paths," *IEEE Trans. Commun.*, vol. 38, pp. 1693–1698, Oct. 1990.
[8] V. A. Margo, D. M. Etter, N. C. Carlson, and J. H. Gross, "Multiple short-length adaptive filters for time-varying echo cancellation," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, vol. I, Minneapolis, MN, Apr. 1993, pp. 161–164.
[9] J. Homer, L. M. Y. Marcels, R. R. Bitmead, B. Wahlberg, and F. Gustafsson, "LMS estimation via structural detection," *IEEE Trans. Signal Processing*, vol. 46, pp. 2651–2663, Oct. 1998.
[10] J. Homer, "Detection guided NLMS estimation of sparsely parametrized channels," *IEEE Trans. Circuits Syst. II*, vol. 47, pp. 1437–1442, Dec. 2000.
[11] D. L. Duttweiler, "Proportionate normalized least-mean-squares adaptation in echo cancelers," *IEEE Trans. Speech, Audio Processing*, vol. 8, pp. 508–518, Sept. 2000.
[12] J. Benesty and S. L. Gay, "An improved PNLMS algorithm," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, vol. II, Orlando, FL, May 2002, pp. 1881–1884.
[13] G. Strang and T. Nguyen, *Wavelets and Filter Banks*. Cambridge, MA: Wellesley-Cambridge , 1996.
[14] S. Hosur and A. H. Tewfik, "Wavelet transform domain adaptive FIR filtering," *IEEE Trans. Signal Processing*, vol. 45, p. 617-130, March 1997.
[15] M. Doroslovacki and H. Fan, "On-line identification of echo-path impulse responses by Haar-wavelet-based adaptive filter," in *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, Detroit, MI, May 1995, pp. 1065–1068.
[16] S. D. Blunt and K. C. Ho, "Novel sparse adaptive algorithm in the Haar transform domain," in *Proc. IEEE Conf. Acoustics, Speech, Signal Processing*, vol. I, Istanbul, May 2000, pp. 452–455.
[17] "Digital Network Echo Cancellers,", ITU-T Recommendation G.168, 2000.
[18] N. J. Bershad and L. Z. Qu, "On the probability density function of the LMS adaptive filter weights," *IEEE Trans. Acoustics, Speech, Signal Processing*, vol. 37, pp. 43–56, Jan. 1989.
[19] T. K. Moon and W. C. Stirling, *Mathematical Methods and Algorithms for Signal Processing*. Upper Saddle River, NJ: Prentice-Hall, 2000.
[20] K. C. Ho and S. D. Blunt, "Enhanced adaptive sparse algorithms using the Haar wavelet," in *Proc. IEEE Symp. Circuits and Systems*, vol. III, Phoenix, AZ, May 2002, pp. 45–48.

K. C. Ho (S'89–M'91–SM'00) was born in Hong Kong. He received the B.Sc. degree (with First Class Honors) in electronics and the Ph.D. degree in electronic engineering from the Chinese University of Hong Kong, in 1988 and 1991, respectively.

From 1991 to 1994, he was a Research Associate in the Department of Electrical and Computer Engineering at the Royal Military College of Canada, Kingston, ON, Canada. From January 1995 to August 1996, he was a Scientific Staff Member at Bell-Northern Research, Montreal, QC, Canada. From September 1996 to August 1997, he was a Faculty Member in the Department of Electrical Engineering at the University of Saskatchewan, Saskatchewan SK, Canada. Since September 1997, he has been with the University of Missouri-Columbia, where he is currently an Assistant Professor in the Electrical and Computer Engineering Department. He is also an Adjunct Associate Professor at the Royal Military College of Canada. Since 1995, he has been active in the development of the ITU recommendation G.168: Digital Network Echo Cancellers. He has received three patents from the United States in the area of telecommunications. His research interests are in statistical signal processing, wireless communications, source localization, wavelet transform and the development efficient adaptive signal processing algorithms for various applications including echo cancellation, equalization, time delay estimation and system identification. He is currently the Editor of *ITU Recommendation G.168*.

Dr. Ho was the recipient of the Croucher Foundation Studentship from 1988 to 1991.

Shannon D. Blunt (S'96–M'02) was born in Memphis, TN. He received the B.S. (*cum laude*), M.S, and Ph.D. degrees in electrical engineering from the University of Missouri-Columbia (MU), in 1999, 2000, and 2002, respectively.

Currently, he is with the Radar Division, U.S. Naval Research Laboratory, Washington, DC. His research interests include adaptive signal processing, wavelets, system identification, statistical signal processing, and estimation theory.

Dr. Blunt was awarded the Donald K. Anderson Graduate Student Teaching Award at MU in electrical engineering in 2000 and the MU Outstanding Graduate Student award in Electrical Engineering in 2001. He is a member of Eta Kappa Nu, Tau Beta Pi, and Sigma Xi.