Adaptive Frequency Sampling Filters

ROBERT R. BITMEAD, MEMBER, IEEE, AND BRIAN D. O. ANDERSON, FELLOW, IEEE

Abstract—We present two new structures for adaptive filters based on the idea of frequency sampling filters and gradient based estimation algorithms. These filters have a finite impulse response (FIR) and can be thought of as attempting to approximate a desired frequency response at given points on the unit circle. The filters operate in real time with no batch processing of signals as is the case when using the discrete Fourier transform. They result in a marked reduction in dimension of the time-domain problem of fitting an Nth-order FIR transversal filter to a collection of length N transversal filters and further to a collection of N scalar filters. The advantages of this are then discussed.

I. INTRODUCTION

ADAPTIVE FILTERS find application in many areas of signal processing including automatic equalization, adaptive antenna arrays, doppler radar systems, adaptive line enhancers, noise and echo cancellers, adaptive control, pattern recognition, etc. One particularly familiar and useful approach to the construction of adaptive filters is to choose a suitably parameterized model structure to approximate the dynamic system under investigation and then to select an algorithm to tune or adapt the variable parameters of this model to improve the approximation, possibly in a time-varying environment, as measured by some given criterion. The thrust of the work of this paper will be to present a modeling structure together with appropriate adaption algorithms which for certain classes of problems results in better, more predictable adaptive performance than current schemes.

Perhaps the most familiar structure for an adaptive filter is the tapped delay line or transversal filter, and we shall naturally refer to it as the yardstick for comparison with our structure to be presented. We shall briefly describe the adaptive tapped delay line filter—its implementation is illustrated in Fig. 1.

Suppose we are given two random sequences \( \{x_i\} \) and \( \{d_i\} \) and we wish to approximate \( \{d_i\} \) as closely as possible by an Nth-order moving average of the \( x_i \). Then, denoting \( X_i^T = (x_{iN}, x_{iN-1}, \ldots, x_{i1}) \) and \( W^T = (w_1, w_2, \ldots, w_N) \), we seek a value of \( W \) to minimize the criterion \( E(d_i - W^T X_i)^2 \). This is a familiar Wiener filtering or linear regression problem for which solution procedures have long been studied in certain cases, e.g., when the two sequences are wide sense stationary and covariances are known. In the slowly time-varying situation with unknown distributions different procedures to estimate \( W \) are used as described below.

There are two immediate points to note about this filter. Firstly, it is finite impulse response (FIR) and indeed the optimum solution for \( W \) can be regarded as a best FIR approximation of length \( N \) to the system linking \( \{x_i\} \) and \( \{d_i\} \). Since it is FIR there are no problems of filter stability. Secondly, the performance criterion above is quadratic in \( W \) so that simple gradient-based methods for estimating the minimising \( W \) can be used. One common gradient-based estimation algorithm is LMS [1]

\[
W_{i+1} = W_i + \mu X_i (d_i - X_i^T W_i)
\]

where \( \mu \) is a fixed gain which determines the convergence rate, and hence also the ability to track time variations.

Transversal adaptive filters, with the above structure and algorithm, may be viewed as attempting to find the best FIR approximation to a desired response by directly estimating the values \( \{W_i\} \) of the impulse response. These filters then have an obvious "time-domain character" about them and this is entirely suited to many applications such as echo cancellers where the resulting impulse response has heuristic meaning. There are other applications however, such as discussed by Griffiths [2], where the impulse response is not of primary interest but rather the frequency response is desired. For this class of problems, where frequency-domain information is desired, it may be more sensible to attempt the adaptation of the filter in a frequency-domain setting. It is this kind of approach to adaptive filtering which we present.

Rather than trying to fit an FIR system to that desired, we approach the problem by estimating the value of the desired frequency response at specified points around the unit circle. In doing this we are freely admitting that we are making an approximation to the ideal system. The course taken is to use fixed frequency sampling filters (FSF) with adaptive gains.

These frequency sampling filters are familiar from FIR digital filter design and will be more fully discussed in Section II. They effectively act as a bank of comb filters,
each passing a single narrow band. We process the complex-valued gain of each band independently and, when tuned, the gains approximate the value of the transfer function across each of the bands. This has the effect of breaking down the adaptation problem from a possibly very large dimensional quadratic minimization to a collection of two-variable quadratic minimizations—the two variables being the real and imaginary (or in-phase and quadrature) parts of the frequency response at the given frequency. We then show how this 2-D problem can be decoupled into two scalar minimizations by proper choice of adaptation algorithm. This yields a potentially very fast and predictable convergence rate.

While the idea of fitting wide sense stationary time series in the frequency domain is certainly not new and the use of frequency domain adaptive filters has been proposed and examined by others, see, e.g., [3]–[7], [16], one novelty of our approach is the use of FSF’s as opposed to using the fast Fourier transform (FFT) for the separation of frequency components. The main difference is that the frequency sampling filters are isochronous with the signal sampling frequency; that is, the FSF produces a real-time output sequence as a digital filter with clock frequency identical to and in step with the input sampling frequency. This is to be contrasted with the FFT which requires batch processing of N time samples of the signal at a time to produce spectrum estimates at discrete frequencies spaced by N−1 times the sampling frequency. This real-time isochronous operation of the FSF’s had led to their output being described as a “sliding spectrum” [8]. Comparing this with the FFT it becomes apparent that the isochronous operation of each of the FSF’s on the continuous input data stream implements a filter which provides one frequency element of the DFT of the previous N samples.

There are a number of implications:

1) As noted above, there is a potential for speeding up the convergence relative to the usual time-domain LMS algorithm (though it may well be that the applications context makes it somewhat pointless to make such a comparison). But what of the comparison between FSF and FFT based approaches? Asymptotically, one would expect that, with comparable settings of gain and with the same frequencies appearing in the samples of the frequency response, the two approaches would be the same. However, as illustrated by simulation the transient behavior of the FSF is clearly faster than that of an FFT method, where nothing happens for N time samples. While the FFT speed may be acceptable in some signal processing schemes, it is far less likely to be satisfactory in an adaptive control scheme, where estimates of the plant are used to tune the controller: with the scheme of this paper, rapid (in the transient phase) adaption of the plant is possible, as needed.

2) Unless one implements more than one FFT scheme at the same time, the FFT scheme produce estimates of the frequency response at frequencies which are evenly spaced. In contrast, at the cost of very little additional complexity, the FSF scheme is not as restricted. With an eye to control applications, one could conceive of situations where equal spacing of the logarithm of the sample frequencies was desired, and an FSF might be more suited to this. In such a case, the additional FSF complication only involves the addition of further parallel processing channels which do not interact with those to which they are added.

3) No problems with circular convolution will arise. Circular convolutions, rather than linear convolutions, can occur in at least two situations where finite length time-domain data is being convolved: when one does a time-domain convolution and deliberately makes the data periodic, and when one uses an FFT to do the convolution without any of the standard precautions (appending of input zeros, or using overlap-save or overlap-add processing). The circular convolution difficulty, which is equally pressing when a FIR filter is being excited by an input stream, is noted in [3] and [6] but not mentioned in [5] and [7]. It is dealt with in [16]. In the scheme of this paper, convolutions are performed by passing actual signals into an actual digital filter with FIR, i.e., convolutions are not evaluated using a computer program where there is the risk or probability of replacing a finite impulse response by a periodic one, nor are they evaluated via an FFT.

4) The use of a sliding spectrum allows us to implement a simple real-time adaptation scheme which attempts to minimize a collection of independent scalar unconstrained quadratic problems yielding an FIR filter with real coefficients. FFT methods such as [5]–[7] have this decoupling property; however, as explicitly demonstrated by Waltzman and Schwartz [3], they require a constrained vector minimization if a filter with real coefficients is desired.

5) There will be a difference in the computational (and hardware) requirements between FSF, FFT, and LMS time-domain algorithms. Unless one is using FFT algorithms simply to speed up the estimation of an impulse response, the LMS algorithm cannot easily be compared to the other two algorithms, since there is no direct relation between the length of an impulse response (one of the factors governing the LMS time-domain algorithm complexity) and the fine structure of the associated frequency response. Also, it can be hard to compare FSF and FFT complexities, unless the frequency response is estimated at the same frequencies for each algorithm and as argued above, this may not be the case. Nevertheless, if these frequencies are the same, an FFT scheme can be expected to offer computational savings. Since the hardware may well be different, however, the comparison may not be meaningful. Moreover, in the FSF scheme, the processing is distributed between various channels, which means that whereas the overall processing rate may be high, the rate per channel may be very moderate.

Having said all this, we would also emphasize that the convergence rate analysis of the algorithm of this paper appears to rest on far fewer heuristic arguments than those employed in most competing algorithms, and in this lies some of the novelty of the paper.

The paper is divided into four sections. In Section II we discuss frequency sampling filters, presenting their background and description as well as two possible implementations of FSF which require only real signals. Section III is devoted to adaptive FSF where we present the algo-
rithms and discuss their performance both theoretically and as indicated by simulations. We conclude with Section IV where we summarize the method’s advantages and drawbacks.

II. FREQUENCY SAMPLING FILTERS

We now briefly describe frequency sampling filters (FSF) and then give some possible implementations of them which require real signals only. FSF are well known in the area of FIR digital filter design [8], [9].

Background and Description

The underlying idea of FSF is that an FIR filter of impulse response length N can be completely described by giving the value of its frequency response at N points equally spaced around the unit circle—this corresponds to specifying the DFT of the filter’s impulse response.

Having chosen the N points we then build filters which have zeros at exactly N - 1 of these points. Such a filter has transfer function

\[ \hat{H}_k(z) = \frac{1 - z^{-N}}{1 - W_N^k z^{-1}} = 1 + W_N^k z^{-1} + W_N^{2k} z^{-2} + \ldots + W_N^{N-1} z^{-N+1} \]

where

\[ W_N^k = \exp\left(\frac{2\pi k}{N}\right) \]

which has zeros at \( W_N^i \) for \( i \neq k \) and may be considered as having a frequency response which passes only those frequencies centered at and close to \( W_N^k \) and excluding others outside this band. As N is made large the frequency response \( \hat{H}_k(e^{j\omega}) \) assumes the form of a sinc-or sampling-function in \( \omega \)—hence the name frequency sampling filter. Banks of these filters centered on different frequencies and with different gains may then be connected in parallel to approximate a desired response. This is further discussed in [8], [9]. It should be noted that the transfer function values are exactly matched at the N points.

One immediate benefit of these FSF can be seen in that the output of each elemental filter \( \hat{H}_k(z) \) is approximately the component of the input sequence at frequency \( \nu_k \) bank of these filters then produces a collection of spectral components which approximate those of the input. These components do overlap slightly in frequency but have zero contribution from other filters at the center frequency. Furthermore, the output spectral components are isochronous with the input, i.e., they arise in real time as a sequence at the same sampling frequency as the input, and hence have been termed a “sliding spectrum” as opposed to batch processed FFT’s. This is similar in concept to the idea of comb filters in radar systems [10].

Equation (2) gives an indication of two possible methods for the implementation of \( \hat{H}_k(z) \). The first method is to construct \( \hat{H}_k(z) \) as an \((N-1)\)-length FIR filter via, say, a tapped delay line. For a bank of these elemental filters \( \{\hat{H}_k(z)\} \) this requires a tapped delay line for each filter. The second approach is to pass the entire input signal through the common “comb” filter \( 1 - z^{-N} \) and then in each of the parallel branches to implement a single pole oscillator filter \((1 - aW_N^k z^{-1})^{-1}\), where \( a \) is a number marginally less than one to guarantee the stability of the oscillators; usually \( a = 1 - 2^{-26} \). Each of these approaches has its obvious advantages.

In neither approach is it envisaged that a convolution is performed by using DFT or FFT methods. Rather an “actual” signal is passed into an “actual” FIR filter; there is accordingly no circular convolution involved.

Implementations for Real–Rational Transfer Function and Real Signals

For transfer function approximation it is necessary to follow each elemental filter with a complex gain equal to the desired value of the frequency response at the particular frequency point on the unit circle. For implementation of these filters for real systems and signals it is desirable to use real operations on real sequences only, and we now show two ways of achieving this. Both methods arise through the collecting together of complex-conjugate pairs of elemental filters.

Clearly the FSF which has a real center frequency, \( H_0(z) \), given by

\[ H_0(z) = \frac{1 - z^{-N}}{1 - z^{-1}} \]

and which passes the dc component, requires only a real gain to specify the systems dc performance. Thus the zeroth elemental filter/gain combination has the form \( A_0 \hat{H}_0(z) \) where \( A_0 \) is the real gain.

The \( k \)th elemental filter with its associated complex gain has the form

\[ \left( A_k + jB_k \right) \frac{1 - z^{-N}}{1 - W_N^k z^{-1}} \]

where \( A_k \) and \( B_k \) represent the real and imaginary parts of the transfer function gain at \( z = W_N^k \). The \((N-k)\)th filter elemental filter has center frequency \( W_N^{N-k} \), and the real character of the impulse response implies that the associated complex gain must be \( A_k - jB_k \). Grouping together the conjugate pair we have

\[ \frac{1 - z^{-N}}{1 - W_N^{N-k} z^{-1}}(A + jB) + \frac{1 - z^{-N}}{1 - W_N^k z^{-1}}(A - jB) = \frac{(1 - z^{-N})(C + Dz^{-1})}{1 - 2 \cos\left(\frac{2\pi k}{N}\right) z^{-1} + z^{-2}} \]

(4)
where $C = 2A$ and

$$D = -2A\cos\left(\frac{2\pi k}{N}\right) - 2B\sin\left(\frac{2\pi k}{N}\right)$$

so that the filter may be constructed by a comb filter $1 - z^{-N}$ followed by a second-order (damped) oscillator followed by a two-tap tapped delay line with real weights $C$ and $D$. This is illustrated in Fig. 2.

The second method appears more complicated, involving discrete Hilbert Transformers, but will be shown to possess certain advantages. We recognise that the frequency response of an ideal Hilbert Transformer is given by

$$\mathcal{H}(e^{j\omega}) = \begin{cases} -j, & 0 \leq \omega < \pi \\ j, & \pi < \omega < 2\pi \end{cases}$$

so that the left-hand side of (4) may be written

$$\begin{align*}
\left( \frac{1 - z^{-N}}{1 - W_N^k z^{-1}} + \frac{1 - z^{-N}}{1 - W_N^{-k} z^{-1}} \right) & (A - \mathcal{H}(z)B) \\
& = \left( 1 - z^{-N} \right) \left( 1 - \cos\left(\frac{2\pi k}{N}\right) z^{-1} \right) 2 \\
& \quad \div \left( 1 - 2\cos\left(\frac{2\pi k}{N}\right) z^{-1} \right). \\
& = \left( 1 - z^{-N} \right) \left( 1 - \cos\left(\frac{2\pi k}{N}\right) z^{-1} \right) 2 \\
& \quad \div \left( 1 - 2\cos\left(\frac{2\pi k}{N}\right) z^{-1} \right).
\end{align*}$$

The filter pair admits the realization shown in Fig. 3 using an ideal Hilbert Transformer.

The ideal Hilbert Transformer acts as a 90° phase shifter at all frequencies and is unrealizable as such because its impulse response is noncausal and stretches from $-\infty$ to $\infty$ with odd symmetry. Approximate Hilbert Transformers must therefore be used, the impulse responses of which must approximate the impulse response of the ideal case, left-truncated and right-shifted to become causal. The right shift, in effect the same as a delay, necessitates the incorporation of delays in paths parallel to the transformer. The design of simple wide-band bandpass FIR digital Hilbert Transformers has been discussed by Rabiner and Schafer [11] and they tabulate many minimax design data with specifications of passband (at frequencies close to the real axis Hilbert Transformation corresponds essentially to producing the quadrature signal from the in-phase component, which is not a well-posed problem), transfer function approximation error and impulse response length/delay (the delay is half the impulse response length). In simulations these designs with impulse response length as low as seven were entirely satisfactory and easily implemented.

We next move on to discuss the adaptive implementation of these filters and their performance with the associated adaptation algorithms.

III. Adaptive Frequency Sampling Filters

Estimation Structures and Adaptive Algorithms

Having specified our chosen modeling structure of a bank of FSF's with pairs of variable real gains, the next step in formulating the adaptive filter is to describe the general estimation structure and devise suitable algorithms for the adaptation of the variable gains in the model.

The general estimation structure is illustrated in Fig. 4 and we shall refer to this for the definition of symbols, signals, etc. Here we have

$$H_k(z) = \begin{cases} 
\frac{(1 - z^{-N})}{1 - 2\cos\left(\frac{2\pi k}{N}\right) z^{-1} + z^{-2}}, & k \neq 0 \\
(1 - z^{-N}) \left( 1 - \cos\left(\frac{2\pi k}{N}\right) z^{-1} \right) 2 \\
1 - 2\cos\left(\frac{2\pi k}{N}\right) z^{-1} + z^{-2}, & k = 0
\end{cases}$$

for our collection of real valued FSF's, the alternative depending on whether or not Hilbert Transformers are used, i.e., whether the setup of Fig. 2 or Fig. 3 applies. It should be noted that the hypothetical "Ideal System" $G(z)$ in Fig. 4 with input sequence $\{x_i\}$ and output sequence $\{y_i\}$ need not be a tangible physical system on which we are conducting our experiment but could simply be a
The aim of the exercise is to choose the variable gains \( A_k, B_k \) so that the error between narrow-band components \( y_{k,t} - \hat{y}_{k,t} \) is small in a certain sense. We treat this as a collection of independent (decoupled) estimation problems, for each of the two-vectors \( (A_k, B_k)^T \) and the scalar \( A_0 \), and concentrate without loss of generality on one real elemental filter from the bank of FSF. We shall discuss the dc filter \( H_0(z) \) later. The rationale behind this decoupling is that the outputs of the FSF for different \( k \) have an approximate orthogonality property by virtue of their being in different narrow frequency bands.

We shall consider firstly the adaptive implementation of the FSF's using the filters of Fig. 2. We call this the direct implementation. We then examine the set up with filters of Fig. 3 involving the Hilbert Transformer and refer to this as the transformed implementation. As we are studying a single representative component, we shall drop the subscript \( k \) from the sequences.

For the direct implementation we recall (4) and write \( \{s_i\} \) for the output of the filter
\[
(1 - z^{-N})(1 - 2 \cos \frac{2\pi k}{N} z^{-1} + z^{-2})^{-1}
\]
driven by \( \{x_i\} \). Thus we seek to have
\[
\hat{y}_i = C s_i + D s_{i-1}
\]
which is equivalent to a simple two-tap delay line. We may use LMS, (1), to update estimates \( \hat{C}_i, \hat{D}_i \) of the optima \( C \) and \( D \) by using \( W_i = (\hat{C}_i, \hat{D}_i)^T \), \( d_i = y_i \) and \( X_i = (s_i, s_{i-1})^T \). This implementation is shown in Fig. 5. The most striking comparison between this adaptive filter and the time-domain transversal filter is that they both utilize the same update algorithm but the time-domain N-vector adaptation is reduced to a collection of \( \lfloor N/2 \rfloor \) two-vector adaptations plus one scalar adaptation. This and its consequences will be discussed more fully after the following presentation of the transformed implementation.

Denoting by \( \{u_i\} \) the output sequence of the FSF
\[
(1 - z^{-N})(1 - 2 \cos \frac{2\pi k}{N} z^{-1} + z^{-2})^{-1}
\]
driven by \( \{x_i\} \) and writing \( \{v_i\} \) as the Hilbert Transform of the sequence \( \{u_i\} \) we have from (5) that ideally
\[
\hat{y}_i = A u_i - B v_i
\]
where \( A \) and \( B \) are, respectively, the real and imaginary parts of the gain at the particular frequency. This again conforms to a two-vector identification problem amenable to the LMS algorithm with \( W_i = (\hat{A}_i, \hat{B}_i)^T \), \( u_i = y_i \) and \( X_i = (u_i, v_i)^T \) in (1).

As remarked earlier an ideal Hilbert Transformer is an unrealizable filter and approximate transformers require the introduction of a delay (usually half the length of the impulse response for FIR approximations) to try to compensate for the noncausality of the ideal filter. Denoting this delay by \( l \) we see that the implementation requires the use of \( \hat{y}_{i-l} = A u_{i-l} - B v_i \). The transformed implementation is shown in Fig. 6.

**Performance and Comparison of Structures**

We now turn to consider the performance of and comparisons between the direct and transformed FSF approaches and the time-domain approach. As stated above the obvious point of comparison between these approaches is the reduction in the dimension of the \( X \) vector in the LMS algorithm. The primary effects of this are to alter the convergence rate of the adaptive filter and to allow us to predict this rate more accurately.

When the sequences \( \{X_i\} \) and \( \{d_i\} \) are random it is straightforward to show that, subject to independence, the LMS algorithm converges exponentially fast to a neighborhood of the best solution at a rate \((1-\beta)^l\), where \( \beta \) depends roughly linearly on \( \mu \) (the adaptation gain) and the minimum eigenvalue of the covariance \( R = E(XX^T) \) or the time average of this quantity, and the degree of dependence of neighboring \( X_i \). This property arises basically because these methods are a form of steepest descent procedure to minimize a quadratic form with average weighting \( R \) with step size proportional to \( \mu \) (which in turn must be less than the reciprocal of the maximum eigenvalue of \( R \) for convergence of LMS). These steepest descent or gradient methods typically exhibit slow convergence when the condition number of the quadratic form is large [13]. This dependence of convergence rate upon condition number of steepest descent methods is alleviated in the use of Newton–Raphson algorithms by premultiplying the gradient step by \( R^{-1} \). This latter technique is exemplified in the more rapidly converging Recursive Least Squares algorithms.

When the dimension of \( X_i \) is large it is often difficult to estimate both the minimum and maximum eigenvalues of
and so a slow convergence rate often occurs because one naturally chooses a conservative value of $\mu$. Quite apart from these problems of ignorance, ill-conditioning problems become increasingly prevalent with increasing dimension. As the dimension is decreased the severity of these problems decreases until, with a small quadratic minimisation, the steepest descent technique is equivalent to Newton-Raphson. The possible advantages of a collection of small dimension minimisations over one large dimension problem then become apparent and we will demonstrate how the proposed implementations allow effective reduction to a collection of $N$ scalar adaptations.

These sorts of arguments are also advanced in [3]–[7] in support of the use of FFT-based adaption algorithms. However, as [3] makes clear, one may well be facing a constrained optimization problem, a fact which can negate much of the advantage.

The dc FSF requires only a single real gain so that its associated adaptive algorithm is a scalar minimization for both FSF implementations. Consequently, the steepest descent methods converge as fast as Newton-Raphson. Furthermore, the convergence rate is easily related to the magnitude of the scalar $\{X_i\}$ sequence. We may therefore easily choose an appropriate value for $\mu$. This will be discussed again later.

We now estimate the convergence rate and performance of the two-vector adaptations of the two adaptive FSF implementations.

As two preliminary comments, we note that when adaptive control is envisaged, the delay associated with use of the second FSF implementation (involving Hilbert transformers) may make this unacceptable. Second, we note that we shall discuss convergence of real and imaginary parts in the following, although for some purposes, convergence of magnitude and phase may be more relevant to consider, and the conclusions may be a little different.

We use the following notation: $C$ and $D$, and $A$ and $B$ are the optimum values of the coefficients (assumed stationary for the moment) for the two implementations with a minimum mean square error criterion; $\{n_i\}$ is the minimum mean square error $y_i - C s_i - D s_{i-1}$ and $y_i - A u_i + B v_i$ (delays assumed to be already accounted for); $\bar{\omega}_i$ is the coefficient estimate error $(C_i - C, D_i - D)^T$ or $(A_i - A, B_i - B)^T$.

The adaptation algorithms for the direct implementation are described by the equations

$$\bar{\omega}_{i+1} = \begin{bmatrix} 1 - \mu \hat{s}_i^2 & -\mu \hat{s}_i s_{i-1} \\ -\mu \hat{s}_i s_{i-1} & 1 - \mu \hat{s}_{i-1}^2 \end{bmatrix} \bar{\omega}_i + \mu n_i \begin{bmatrix} \hat{s}_i \\ s_{i-1} \end{bmatrix}. \tag{7}$$

The associated homogeneous equation for sufficiently small $\mu$ is known to be exponentially convergent under various mild conditions, see, e.g., [12]. Its time constants determine for (7) the time constant associated with convergence of the moments of $\bar{\omega}_i$. We shall now examine the homogeneous version of (7) to gain insight into what these time constants are.

Our arguments lack full rigor, but are certainly suggestive. Because $s_i$ is a narrow-band process, we represent it as

$$s_i = r_i \cos \left( \frac{2\pi k i}{N} + \phi_i \right). \tag{8}$$

Here, $r_i$ and $\phi_i$ are slowly varying processes. The coefficient matrix in (7) can be written as

$$\begin{bmatrix} 1 - \mu \hat{s}_i^2 & -\mu \hat{s}_i s_{i-1} \\ -\mu \hat{s}_i s_{i-1} & 1 - \mu \hat{s}_{i-1}^2 \end{bmatrix} \begin{bmatrix} \mu (\hat{s}_i^2 - s_{i-1}^2) \\ -\mu (s_i^2 - s_{i-1}^2) \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}$$

where

$$\begin{align*}
(s_i + s_{i-1})^2 &= 4r_i^2 \cos^2 \left( \frac{2\pi k i}{N} + \frac{\pi k}{N} + \frac{\phi_i + \phi_{i-1}}{2} \right) \\
(s_i - s_{i-1})^2 &= 4r_i^2 \sin^2 \left( \frac{2\pi k i}{N} + \frac{\pi k}{N} + \frac{\phi_i + \phi_{i-1}}{2} \right) \\
(s_i^2 - s_{i-1}^2) &= r_i^2 \sin 2 \left( \frac{2\pi k i}{N} + \frac{\pi k}{N} + \frac{\phi_i + \phi_{i-1}}{2} \right) \\
&= r_i^2 \sin \left( \frac{2\pi k}{N} + \frac{\phi_i - \phi_{i-1}}{2} \right).
\end{align*}$$

If $\mu$ is small enough in (7), the fast frequency $(4\pi k / N)$ variations in the $s_i, s_{i-1}, s_i s_{i-1}$ terms are inconsequential, i.e., the coefficient matrix can be replaced to a good approximation by its slow-varying part:

$$\begin{bmatrix} 1 - \mu \hat{s}_i^2 & -\mu \hat{s}_i s_{i-1} \\ -\mu \hat{s}_i s_{i-1} & 1 - \mu \hat{s}_{i-1}^2 \end{bmatrix} \approx \begin{bmatrix} 1 - 2\mu r_i^2 \cos^2 \left( \frac{\pi k}{N} \right) & 0 \\ 0 & 1 - 2\mu r_i^2 \sin^2 \left( \frac{\pi k}{N} \right) \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}.$$
We replace \( u, v \) by \( a_i \cos 2\pi ki/N \) and \( \beta_i \sin 2\pi ki/N \), where, if \( u, v \) are Gaussian and exact Hilbert transformers, \( a_i \) and \( \beta_i \) are also Gaussian and band limited to \( 2\pi/N \). The coefficient matrix in (10) becomes

\[
\begin{bmatrix}
1 - \mu a_i^2 \cos^2 \frac{2\pi k_i}{N} & \frac{1}{2} \mu a_i \beta_i \sin \frac{4\pi k_i}{N} \\
\frac{1}{2} \mu a_i \beta_i \sin \frac{4\pi k_i}{N} & 1 - \mu \beta_i^2 \sin^2 \frac{2\pi k_i}{N}
\end{bmatrix}
\]

with the approximation involving elimination of fast time variation. The time constants associated with (10) are then those associated with

\[
\lambda_{i+1} = (1 - \mu a_i^2) \lambda_i \\
\xi_{i+1} = (1 - \mu \beta_i^2) \xi_i
\]

and these are the same, since \( a_i, \beta_i \) are identically distributed.

We may use the results of the Appendix to estimate the time constants of (9a), (9b), (11a), and (11b). These results state that, e.g., (11a) converges with time constant at least as fast as the deterministic equation

\[
\lambda_{i+1} = \left[ 1 - \frac{1}{2} \mu E \left( a_i^2 \right) \right] \lambda_i
\]

when \( \sigma_i \) is ergodic. Generalization to the nonstationary case is possible. The differing time constants of (9) and (11) become evident.

The overall effect of the FSF adaptive filters compared to the time domain is to reduce a single \( N \)-vector problem to a collection of \( N/2 \) two-vector problems plus one scalar problem, with each two-vector problem being equivalent to a pair of one-vector problems.

Whilst pointing out the advantages of these adaptive FSF’s, firstly of the direct implementation over time-domain implementation, and then of the transformed implementation over the direct, it is pertinent to reiterate that the dominant dictator of which filter to use is the applications context. For example, if a measure of transient performance is desired by approximating an impulse response then the use of FSF would clearly be unwarranted over the time-domain approach. Similarly, the use of the transformed implementation instead of the direct implementation is not necessarily better. Indeed the convergence rate of the direct algorithm for the estimate of \( C + D \) may be faster than that of the transformed method—\( C + D \) is related to the transfer function magnitude—so that the direct method is to be preferred in some instances.

**Performance with Noise**

Many calculations relating to the performance of LMS-type algorithms rely upon assumptions of whiteness of the input signals, see [1]. While this assumption can be relaxed [12], it allows us to easily produce good rules of thumb for design and to roughly quantify the filter behavior.

Here, we need to make different sorts of assumptions, using the narrow-band properties of the signals in each channel. We perform an analysis for the transformed algorithm only. Then \( n_i = y_i - Au_i + Bo_i \). In a noiseless situation, with ideal filters with rectangular arbitrarily narrow pass bands, and with perfect Hilbert transformers, we could expect \( n_i \equiv 0 \). However, we shall assume that \( n_i \) is a narrow-band signal representable as

\[
y_i \cos \frac{2\pi k_i}{N} + \delta_i \sin \frac{2\pi k_i}{N}.
\]

Here, \( \gamma_i \) and \( \delta_i \) are slowly varying random process. In order that \( A \) and \( B \) be optimum (least mean square) coefficients, we demand that they minimizes

\[
E \left[ \left( y_i - Au_i + Bo_i \right)^2 \right].
\]

Of course, this ensures that at the optimum, \( E[n_i] = 0 \), \( E[n_i u_i] = 0 \) and \( E[n_i v_i] = 0 \). Let us also suppose that the inexactitudes giving rise to \( n_i \) have a symmetry property which ensures that the probability densities \( p(n_i | u_i) \) and \( p(n_i | v_i) \) are symmetric in \( u_i \) for all \( u, v \), with the associated variances independent of \( u_i \) and \( v_i \). Then \( E[n_i | u_i] = E[n_i | v_i] = 0 \) which implies that \( E[\gamma_i | a_i] = 0 \), \( E[\delta_i | \beta_i] = 0 \), and \( E[\gamma_i^2 | a_i] \) and \( E[\delta_i^2 | \beta_i] \) are independent of \( a_i \) and \( \beta_i \).

Previously, we approximated the homogeneous part of (10), eliminating fast time variations, with (11). If we now include the driving term in (10), the approximation becomes

\[
\begin{align*}
\lambda_{i+1} &= (1 - \mu a_i^2) \lambda_i \\
\xi_{i+1} &= (1 - \mu \beta_i^2) \xi_i
\end{align*}
\]

It is only necessary to study one of these equations.

We now appeal to the slow variation property of \( \sigma_i \) to write (strictly an approximation)

\[
E[\lambda_{i+1} | a_i] = E[\lambda_i | a_i].
\]

(The signal \( \sigma_i \) has a bandwidth of \( 2\pi/N \). Thus if the sampling rate for the parameter update equation is significantly faster than this—as it will have to be in order that sampling occur at a frequency in excess of the overall system bandwidth—the approximation is well justified.)

It follows that

\[
E[\lambda_{i+1} | a_i] = (1 - \frac{1}{2} \mu a_i^2) E(\lambda_i | a_i) - \frac{1}{2} \mu a_i E(\gamma | a_i)
\]

or

\[
E[\lambda_{i+1} | a_i] = (1 - \frac{1}{2} \mu a_i^2) E(\lambda_i | a_i).
\]

Then \( E[\lambda_i | a_i] \to 0 \) with a time constant the same as that of the homogeneous (11a), as then does \( E[\lambda_i] \) and \( E[\lambda | a_i] \).

Next, we look at the second moment. As a preliminary, we consider the quantity \( E[\lambda_i \gamma_i, \sigma_i | a_i] \). From (12a) and the slow variation of \( \gamma_i \) and \( \sigma_i \), we have

\[
E[\lambda_{i+1} \gamma_{i+1} | a_i] = (1 - \frac{1}{2} \mu a_i^2) E[\lambda_i \gamma_i, \sigma_i | a_i] - \frac{1}{2} \mu a_i^2 E[\gamma_i^2]
\]
or, with $E[\gamma_{i+1}]=E[\gamma_i]$ by stationarity,
\[
E[\lambda_{i+1}y_{i+1}\alpha_{i+1}|\alpha_{i+1}]+E[\gamma_i^2]
\]
\[
=(1-\frac{1}{2}\mu\alpha_i^2)(E[\lambda_i\gamma_i\alpha_i]+E[\gamma_i^2])
\]
whence we see that as $i \to \infty$,
\[
E[\lambda_i\gamma_i\alpha_i]=E[\gamma_i^2].
\]
(13)

Now squaring (12a) and conditioning with the slow-variation property gives
\[
E[\lambda_{i+1}^2|\alpha_{i+1}]=\frac{1}{2}(1-\frac{1}{2}\mu\alpha_i^2)E[\lambda_i^2|\alpha_i]+\frac{1}{2}\mu^2\alpha_i^2E[\gamma_i^2]
\]
\[
-\mu(1-\frac{1}{2}\mu\alpha_i^2)E[\lambda_i\gamma_i\alpha_i|\alpha_i]
\]
\[
=(1-\frac{1}{2}\mu\alpha_i^2)^2E[\lambda_i^2|\alpha_i]+\mu(1-\frac{1}{2}\mu\alpha_i^2)E[\gamma_i^2]
\]
by (13). It then follows that
\[
E[\lambda_{i+1}^2|\alpha_{i+1}]=\frac{1}{2}(1-\frac{1}{2}\mu\alpha_i^2)\left(E[\lambda_i^2|\alpha_i]-E[\gamma_i^2]\right)E[\gamma_i^2]
\]
\[
+\frac{1}{2}\mu^2\alpha_i^2\frac{E[\gamma_i^2]}{\alpha_i^2}.
\]

Multiplying by $\alpha_i^2$ and invoking slow variation of $\alpha_i$ and stationarity of $E[\gamma_i^2]$ we have
\[
\alpha_{i+1}^2E[\lambda_{i+1}^2|\alpha_{i+1}]=E[\gamma_i^2]
\]
\[
=(1-\frac{1}{2}\mu\alpha_i^2)^2E[\lambda_i^2|\alpha_i]-E[\gamma_i^2]
\]
Hence [14], provided $E[\alpha_i^2]<2\mu^{-1}$, as $i \to \infty$
\[
\alpha_i^2E[\lambda_i^2|\alpha_i]=E[\gamma_i^2] \text{ a.s.}
\]
and taking expectations
\[
E[\alpha_i^2\lambda_i^2]=E[\gamma_i^2].
\]
(14)

Thus we see that, despite the correlation of all the signals involved in the LMS algorithm, the output error, $\alpha_i\lambda_i$, converges to the optimum value as measured by the first two moments of the error distribution. The correlation of the signals has not affected the performance of the LMS algorithm in this application.

The crucial assumptions in the above analysis have been the slow variation of the magnitude signals such as $\alpha_i$, and the slow variation of the parameter estimate errors $\lambda_i$ relative to the signal frequencies. It is to be expected that as these restrictions become less valid, either by moving to higher frequencies or by increasing the value of $\mu$, the assumptions behind the analysis become less justifiable implying that a change in performance will arise.

IV. Simulation Results

Simulations of both types of adaptive frequency sampling filter were conducted for a variety of different systems including both those FIR systems which could be exactly modeled by a collection of weighted FSFs and systems (FIR and IIR) which could not. The performance of both adaptive filter implementations were compared for both types of systems above, with and without noise, and with differing gains.

The Hilbert transformers for the transformed implementation were approximated by FIR Minimax approximations of [11]. It was found that approximate transformers of impulse response length 7 were workable for frequency separations of $2\pi/20$ radians and that those with length 15 were entirely adequate.

The gain $\mu$ used in both implementations was appropriately scaled to the signal power, as is common, to allow comparison between the convergence rates and it was found that, as predicted, the convergence rate of the transformed implementation was indeed faster than that of the direct implementation especially close to the real axis. The difference between rates decreased as the frequency approached $\pi/2$ radians. Except for systems with transfer functions such as $1-z^{-1}$, which are perfectly describable by the direct implementation, the transformed algorithm appeared to perform more accurately.

We present below several simulation graphs which illustrate for the transformed algorithm the effect of varying the signal to noise ratio and the adaptive gain $\mu$. Similar effects are exhibited by the other implementation. These simulations were carried out with random input sequences to the adaptive filter and measurement noise. An ensemble of 100 simulations was performed and the ensemble mean parameter value together with the ensemble standard deviation are plotted.

There are several effects limiting the accuracy of the transformed adaptive FSF. Firstly, there is model mismatch due to the inability of describing exactly the plant system by a bank of weighted FSFs. This manifests itself as an extra forcing function to the homogeneous parameter update equations resulting in a time-variation of the parameter about the optimum value. The smaller the value $\mu$ is, the smaller is the amplitude of this time variation. Allied with this mismatch and contributing time variations possibly with dc offset is the inaccuracy of the Hilbert transformer. In the simulations presented this latter effect was negligible, however. Associated with model mismatch is the overlapping of neighboring frequency sampling filters due to their nonideal bandpass nature—this too may introduce a time variation and offset in the parameter estimate.

The above sources of parameter error arise even when the adaptive filter operates in the absence of measurement noise. The situation is illustrated in Fig. 7 for the system

\[
G(z) = \frac{1+3z^{-1}+3z^{-2}+z^{-3}}{6+z^{-2}}
\]
at the frequency 0.1π with gain μ = 0.1 with normalization to signal level. The straight lines on the graph represent the real and imaginary parts of G[exp(0.1jπ)]. Note the width of the standard deviation band and the convergence rate.

With the addition of measurement noise at 10 dB below the signal power, the simulation was repeated. The measurement noise added onto the system output was white Gaussian noise independent of the system output. Fig. 8 illustrates the outcome of this simulation. Notice that the predictable effect of the inclusion of this noise is to leave the mean value unaltered while causing a spreading of the standard deviation from the mean due to the included uncorrelated random process. The convergence rate has not been perceptibly altered, although it is known (see Appendix) that an extremely large noise signal can prevent convergence of the filter.

Fig. 9 demonstrates the effect of altering the gain μ. In Fig. 9 μ has been changed from 0.1 to 0.25 (after signal strength scaling) and the noise has been left at 10 dB below the signal. The most startling effects to be noticed are the increased convergence rate, the increased standard deviation and the lack of smoothness.

V. CONCLUSIONS

We have presented two new adaptive filters based on frequency sampling/sliding spectrum ideas and LMS-type estimation algorithms. The first of these filters—the direct implementation—involves a bank of filters each followed by a separate two-tap delay line which incorporates the adaptive gains. This was shown to reduce the N-vector transversal filter problem of the time domain to a collection of [N/2] two-vector transversal filter problems plus a single scalar problem. The second filter—the transformed implementation—involves a similar bank of filters and the use of a Hilbert Transformer at the output of each together with the two adaptive gains. While this latter scheme is more complicated in concept, it allows us to further decouple the two-vector adaptations so that the single N-vector problem reduces to N scalar ones.

The advantages of this reduction in dimension of the adaptation were shown to be related to the convergence rate (especially the “transient” convergence rate), to our ability to predict this rate and performance, to have confidence in our design, of parameters for different frequency bands. The convergence rate is related to the eigenvalue properties of a certain matrix having the same dimension as the transversal filter involved. By lowering this dimension through using FSF methods we reduce the likelihood of condition number problems and increase the predictability of the adaptive performance.

Finally, it should again be remarked that we do not intend that adaptive FSF’s should be used to the exclusion of familiar time domain filters, especially tapped delay line filters, but rather that they should be thought of as another valid alternative possessing certain advantages in many situations. The choice of adaptive filter should be dictated by the ultimate application.

APPENDIX

We consider the apparently trivial problem of the stability of the equation

\[ x_{k+1} = a_k x_k, \quad x_1 = 1. \]  (A1)

The problem is made nontrivial by the assumption that \( \{a_k\} \) is an ergodic random process. We shall find a sufficient condition for the exponential stability of \( \{x_k\} \), i.e., for the existence of a constant \( \alpha \in (0, 1) \) such that \( \alpha^{-k} x_k \to 0 \) a.s., or equivalently, the existence of \( \beta(\omega) \) and constant \( \alpha \in (0, 1) \) such that \( |x_k| \leq \beta(\omega) \alpha^k \). A similar result has been obtained in continuous-time by Parthasarathy and Evan-Iwanowski [14].

It is clear that the solution of (A1) is exponentially convergent if and only if the solution of \( y_{k+1} = |a_k| y_k, \quad y_1 = 1 \) is exponentially convergent.

We need to partition the set of events \( \Omega \) into \( \Omega_0 \cup \Omega_1 \), where \( \Omega_0 = \{\omega|a_k = 0 \text{ for some } k\} \) and \( \Omega_1 = \{\omega|a_k \neq 0 \text{ for any } k\} \). We first argue that \( P(\Omega_0) = 0 \). Suppose that \( P(\Omega_1) = 0 \) (Note: this is a probability, not a probability density.) Then it is clear that \( P(\Omega_0) = 0 \), \( \Omega_0 \) being a countable union of events of zero probability. Suppose then that \( P(\Omega_1) \neq 0 \). Define \( x_k \) by \( x_0 = 0 \) if \( a_k \neq 0 \), \( x_1 = 1 \) if \( a_1 = 0 \). By the ergodic theorem

\[ \lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} x_k \to E[x_1] = P(\alpha_1 = 0) \]

w.p.l. Thus if \( P(\alpha_1 = 0) \neq 0 \), we have w.p.l an infinite
number of $x_i$ are 1, and thus an infinite number of $\alpha_i$ are zero, i.e., $P(\Omega_0) = 1$.

If $P(\Omega_0) = 1$, exponential stability is immediate. So assume $P(\Omega_1) = 1$. On $\Omega_1$, we shall study the solution of the related equation

$$y_{k+1} = |\alpha_k| y_k, \quad y_1 = 1$$

which can be written as

$$y_{N+1} = \left( \frac{1}{N} \sum_{k=1}^{N} |\alpha_k| \right)^N y_1.$$  

Now

$$\left( \frac{y_{N+1}}{y_1} \right)^{1/N} = \exp \left( \frac{1}{N} \sum_{k=1}^{N} \ln |\alpha_k| \right),$$

so that

$$\lim_{N \to \infty} \left( \frac{y_{N+1}}{y_1} \right)^{1/N} = \exp \left( E[\ln |\alpha_1|] \right). \quad (A2)$$

It follows that a necessary and sufficient condition for exponential stability is

$$E[\ln |\alpha_1|] < 0 \quad (A3)$$

provided this expected value exists. Evidently, the dependence properties of the $\{\alpha_i\}$ are irrelevant. By Jensen’s inequality, a sufficient condition is

$$\ln E[|\alpha_1|] < 0 \quad (A4)$$

or

$$E[|\alpha_1|] < 1. \quad (A5)$$

If (A5) is satisfied then using events like $\Omega_0$ and $\Omega_k = \{ \omega \mid -\epsilon < \alpha_k < \epsilon \text{ for some } k \}$ it is possible to extend this sufficient condition to hold even when the expectation in (A3) does not exist.

This condition should be compared to the sufficient condition of [14]. A more refined sufficient condition follows by rewriting (A3) as

$$E[\ln |\alpha_1| |\alpha_1 > 0] P(\alpha_1 > 0) + E[\ln (-|\alpha_1|) |\alpha_1 < 0] P(\alpha_1 < 0)$$

which again by Jensen’s inequality yields

$$\ln E[|\alpha_1| ] > 0 \quad (A6)$$

or

$$E[|\alpha_1| ] > 0 \quad (A7)$$

Hölder’s inequality shows that if (A5) holds, then so must (A6). So there may be instances where (A6) can be used to conclude exponential stability, but not (A5). If $\alpha_i$ has a symmetric density, (A5) and (A6) are equivalent.

**Example 1:** Suppose $\alpha_1$ is Gaussian, with mean 0 and variance $\sigma^2$. Then

$$E[|\alpha_1| ] = \sqrt{\frac{2}{\pi}} \sigma \quad \text{and} \quad \sigma < \sqrt{\frac{\pi}{2}}$$

is sufficient for stability.

**Example 2:** Suppose $\alpha_1$ is Gaussian, with mean $\mu$ and variance $\sigma^2$. Then

$$E[|\alpha_1| ] = \sqrt{\frac{2}{\pi}} \sigma \exp(-\frac{\mu^2}{2\sigma^2}) + \mu \text{erf}(\frac{\mu}{\sqrt{2}\sigma}).$$

**Example 3:** $\alpha_1 = 1 - \beta_i^2$ where $\beta_i$ is Gaussian with mean zero and variance $\sigma$. Then

$$E[|\alpha_1| ] = (\sigma^2 - 1) \left( 1 - 2 \text{erf} \left( \frac{1}{\sqrt{2}\sigma} \right) \right) + \frac{4\sigma}{\sqrt{2\pi}} \exp \left( -\frac{1}{2\sigma^2} \right).$$

If $\alpha_i$ is not another ergodic but has a suitable mixing property, such as $\phi$-mixing with a summability condition on the square roots of the mixing parameters, see [15] then we can still obtain a convergence result which may be applied in nonstationary cases. Assume that $P(\Omega_1) = 1$. Then (A3) is replaced by

$$\lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} E[|\alpha_i|] < 0 \quad (A7)$$

and (A5) by

$$\lim_{N \to \infty} \frac{1}{N} \sum_{i=1}^{N} E[|\alpha_i|] < 1. \quad (A8)$$

The application we make of these results is to equations of adaptive estimation which arise with $\alpha_i = 1 - \beta_i^2$, $\beta_i$ ergodic or $\phi$-mixing, for which a sufficient condition is $E[|\alpha_i| ] < 1$.

**References**


Robert R. Bitmead (S’76—M’79) was born in Sydney, Australia, in 1954. He received the B.Sc. degree in applied mathematics in 1976 from the University of Sydney, and the M.E. and Ph.D. degrees in electrical engineering from the University of Newcastle, Australia, in 1977 and 1979, respectively.

During 1979 and early 1980 he held a Post-Doctoral research position in the Department of Electrical Engineering at the University of Newcastle. In 1977 he was awarded a Fulbright grant for a period of study in the Information Systems Laboratory at Stanford University, CA. Since 1980 he has been a faculty member in the Department of Electrical and Electronic Engineering at James Cook University, Townsville, Australia. His current research interests encompass adaptive and learning systems, linear multivariable control, stochastic processes and digital signal processing.

Brian D. O. Anderson (S’62—M’66—SM’74—F’75), for a photograph and biography please see page 510 of this issue.

A Performance Analysis of Adaptive Line Enhancer-Augmented Spectral Detectors

JOHN T. RICKARD, MEMBER, IEEE, JAMES R. ZEIDLER, MEMBER, IEEE, MAURO J. DENTINO, MEMBER, IEEE, AND MARK SHENSA

Abstract—This paper discusses the receiver operating characteristic performance of Adaptive Line Enhancer (ALE) augmented spectral detectors for sinusoidal signals in both stationary white Gaussian noise and in nonstationary noise. The detectors considered are based on the discrete Fourier transform (DFT) and include both cases with and without incoherent integration. Analytical expressions are provided for the detector output probability density functions in the stationary noise case. Extensive Monte Carlo simulation results are used to verify these expressions, and to treat the nonstationary noise case.

I. INTRODUCTION

Several distinct types of adaptive detector structures have been proposed for use in environments where insufficient a priori information on the signal and noise statistics is available to design fixed optimal detectors. Detector structures that utilize adaptive mean level detection algorithms, sequential detection, nonparametric detection, pattern recognition, and decision-directed feedback techniques are discussed in numerous references [1]–[7] for a variety of signal and noise environments. Adaptive detectors are typically ad hoc designs that are based on knowledge of some general features of the signal and noise, but are capable of adjusting their internal structure through observation of the received data. Because their structure is data-dependent, a detection performance analysis is generally much more difficult than for nonadaptive detectors.

In this paper, we are concerned with the general narrow-band-signal-in-broad-band-noise detection problem. The conventional detector structure for this problem is implemented as a bank of narrow-bandpass complex (in-phase and quadrature) filters whose passbands span the frequency range of interest. The complex output of each filter is sampled at the end of a fixed observation interval and quadratically detected to form the basic detection statistic for each passband. This statistic is optimal (in the Neyman–Pearson sense) for detecting a sinusoid of known frequency, centered in a given passband, and unknown phase in white Gaussian noise of known power [26]. The above system may be implemented using the fast discrete Fourier transform (DFT) algorithm to perform the complex filtering operation [9], [10], in which case the detection statistics are the set of magnitude-squared DFT coefficients corresponding to the frequency range of interest. If the narrow-band signal duration exceeds the coherent integration time of the DFT, then the detected filter outputs for a given passband may be incoherently integrated to form an improved statistic [1], [2], [7]–[9], [12].

In recent years, the Adaptive Line Enhancer (ALE) has been proposed for use as a prefilter to the conventional DFT detector described above [11]–[17]. The ALE is an adaptive digital transversal filter that is designed to suppress broad band components in its input, while passing narrow-band components with little attenuation. Extensive analytical treatments of ALE convergence [11], [18]–[20], mean asymptotic weight behavior [11], [18], [21], [22], and output statistics [23]–[25] for particular classes of inputs have previously been published.

0098-4094/81/0600-0534$00.75 ©1981 IEEE