Adaptive Notch Filtering for the Retrieval of Sinusoids in Noise

D. V. BHASKAR RAO AND SUN-YUAN KUNG, SENIOR MEMBER, IEEE

Abstract—In this paper, an adaptive notch filter is developed (employing a frequency domain and time domain analysis) for the enhancement and tracking of sinusoids in additive noise, colored or white. The notch filter is implemented as a constrained infinite impulse response filter with the constraint enforced by a single parameter termed the debiasing parameter. The resulting notch filter requires few parameters, facilitates the formation of the desired band rejection filter response, and also leads to various useful implementations (cascade, parallel). For the adaptation of the filter coefficients, the stochastic Gauss-Newton algorithm is used. The convergence of this updating procedure is established by studying the associated differential equation. Also, it is shown that the structure present in the problem enables truncation of the gradient, thereby reducing the complexity of adapting the filter coefficients. Simulation results are presented to substantiate the analysis, and to demonstrate the potential of the notch filtering technique.

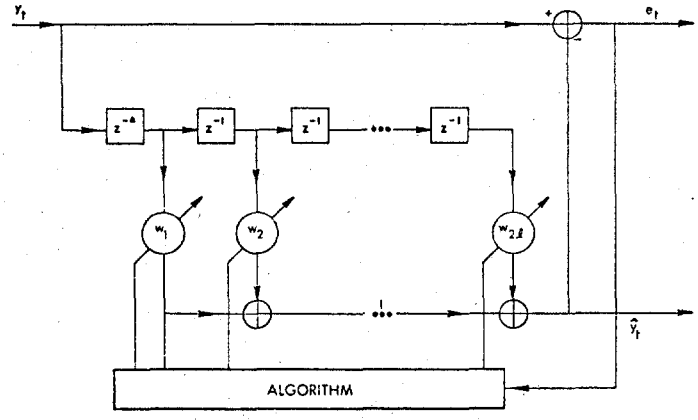

Fig. 1. ALE using a tap delay line filter.

I. INTRODUCTION

THE problem of retrieving sinusoidal/narrow-band signals has received a great deal of attention. The problem is a multifaceted one in that sometimes the frequencies of the sinusoids are of interest, at times an enhanced signal is desired, and sometimes the tracking of a signal with time varying frequency is of interest. In this paper, we concentrate on the problem of line enhancement and tracking. We assume that we are not dealing with extremely short data records, as is often assumed, for the problem of frequency estimation. For this purpose, we develop a filter that falls into a class of filters termed adaptive line enhancers (ALE). Perhaps the most popular ALE is the tapped delay line adaptive filter (see Fig. 1). Because of the recent emergence of other structures for the problem, we refer to this popular adaptive filter as the finite impulse response ALE (FIR-ALE). This filter was first conceived in the now famous paper by Widrow *et al.* [1], and later developed by Griffith [2] and by Treichler [3], [4]. Extensive work has also been carried out to understand its enhancement properties, adaptivity properties, tracking behavior, etc. [5]-[7]. An important advantage of the filter is the simplicity of the adaptive scheme. Also, it cleverly exploits the coherent nature of the signal to separate the narrow-band and broad-band (noise) signals. The decorrelation delay Δ (see Fig. 1), suitably chosen, can ensure that the noise samples have less correlation as opposed to the strongly correlated sinusoidal signal. When the decorrelation delay is set equal to 1, the approach can be considered to be fitting an autoregressive (AR) model.

Although the AR model has been widely used in model based spectral estimation with a great deal of success, it is not without shortcomings. In the presence of noise the model is not the most appropriate one, and as a result, a filter with large length is required to achieve good performance [8]. This large length may outweigh the computational simplicity, leaving room for smaller filters based on more complex models and employing, possibly, more complex adaptive schemes. Smaller filter length implies fewer parameters and also is sometimes accompanied by better numerical behavior, less adaptation noise, etc. Recently, an ALE based on the autoregressive moving average (ARMA) model, the ARMALE, has been proposed for the problem [9]. Although the ARMALE promises to be useful, its generality may be excessive for our purpose.

In this paper, we propose a constrained IIR filter that is specifically designed for this problem. The intuitive notion that a highly tuned filter at the appropriate frequency can be used to retrieve the sinusoidal signals is formalized and viewed in the context of modern model based signal processing. The special filter developed, unlike the other methods, facilitates the formation of a band-rejection filter and can be easily configured to obtain a tuned filter. Furthermore, the parameterization enables the filter to deal with additive noise, white or colored. Moreover, as the signal frequency is unknown, the adaptation of the parameters enables the center frequency to vary and, eventually, converge to the desired location.

The organization of the paper is as follows. We first conduct

Manuscript received October 4, 1982; revised December 5, 1983. This work was supported in part by the U.S. Office of Naval Research under Contract N00014-81-K-0191, by the National Science Foundation under Grant ENG-7908673, and by the U.S. Army Research Office under Grant DAAG29-79-C-0054.

D. V. Bhaskar Rao is with the Department of Applied Mechanics and Engineering Sciences, University of California at San Diego, La Jolla, CA 92093.

S.-Y. Kung is with the Department of Electrical Engineering Systems, University of Southern California, Los Angeles, CA 90089.

0096-3518/84/0800-0791$01.00 © 1984 IEEE

a frequency and time domain study to derive the appropriately parameterized filter. Then we determine a suitable adaptive scheme to estimate the parameters and establish its convergence. Also, it is shown that a truncated gradient can often be used, reducing the computational complexity. Some simulations are presented which indicate that the method is promising, and particularly appropriate, under certain conditions.

II. Frequency Domain Analysis

First, we start with the intention of determining a filter which can give unbiased and accurate estimates of the frequencies, and later show how it lends itself naturally for line enhancement and tracking. A steady-state frequency domain analysis appears to be a natural choice due to the inherent frequency domain structure in the problem, i.e., signals are being discriminated on the basis of their bandwidth. The input signal y_t at any instant of time t is given by

$$y_t = x_t + n_t \tag{1}$$

where x_t denotes the uncorrupted sinusoidal signal, and n_t the additive noise sample at time t. Let $S_y(\omega)$, $S_x(\omega)$, and $S_n(\omega)$ be the power spectrum of y_t, x_t, and n_t, respectively. The noise is assumed to have a smooth spectrum and is independent of the signal. Hence

$$S_y(\omega) = S_x(\omega) + S_n(\omega). \tag{2}$$

$S_x(\omega)$, which represents the line spectrum of p sinusoidal signals, is given by

$$S_x(\omega) = \sum_{1}^{2p} \sigma_i^2 \delta(\omega - \omega_i) \tag{3}$$

where σ_i^2 is the power at frequency ω_i and $\delta(\cdot)$ is the Dirac delta function. On passing the signal y_t through a digital filter $H(z;\phi)$, the output power J is given by

$$J = \frac{1}{2\pi}\int_{-\pi}^{\pi} |H(e^{j\omega})|^2 S_y(\omega)\, d\omega$$

where ϕ is the parameter vector characterizing the filter. For example, for the FIR-ALE the parameter vector $\phi = [w_1, \cdots, w_p]^T$ and

$$H(e^{j\omega};\phi) = 1 + w_1 e^{-j\omega} + \cdots + w_p e^{-j\omega p}.$$

Using (2) and (3), we have

$$J = \frac{1}{2\pi}\int_{-\pi}^{\pi} S_n(\omega)|H(e^{j\omega})|^2\, d\omega + \sum_{1}^{2p} \frac{\sigma_i^2}{2\pi}|H(e^{j\omega_i})|^2$$

$$= A(\phi) + B(\phi) \tag{4}$$

where

$$A(\phi) = \frac{1}{2\pi}\int_{-\pi}^{\pi} S_n(\omega)|H(e^{j\omega})|^2\, d\omega$$

and

$$B(\phi) = \sum_{1}^{2p} \frac{\sigma_i^2}{2\pi}|H(e^{j\omega_i})|^2. \tag{5}$$

Ideally, it is desirable to minimize $B(\phi)$ alone. This would result in the filter having a frequency response with zero response at the correct frequency locations, resulting in unbiased estimates of the frequencies. This is possible only if there was no noise. In the presence of noise, minimizing J results in a compromise situation. $A(\phi)$ also depends on ϕ, and so, unless special steps are taken, it affects the minimization process. Also as $A(\phi)$ is proportional to the noise power, the effect increases with an increase in noise power.

The source of the problem from the above analysis can be clearly seen to be the dependence of the noise power contribution $A(\phi)$ on the parameter vector ϕ. A simple solution is to make $A(\phi)$ constant and thus independent of ϕ. As no *a priori* information regarding the noise is assumed, a sufficient constraint is that the filter must have a flat (unit) power spectrum except at a finite number of notch frequencies, i.e., a filter with a transfer function that is 1 almost everywhere, as shown in Fig. 2. The ideal notch filter (zero bandwidth) is given by

$$H(e^{j\omega};\phi) = \begin{cases} 0, & \omega_i = \omega_1', \cdots, \omega_{2p}' \\ 1, & \text{elsewhere.} \end{cases} \tag{6}$$

Using a filter as in Fig. 2 for $H(z)$, we have

$$J(\phi) = \frac{1}{2\pi}\int_{-\pi}^{\pi} |H(e^{j\omega})|^2 S_n(\omega)\, d\omega + \frac{1}{2\pi}\int_{-\pi}^{\pi} |H(e^{j\omega})|^2 S_x(\omega)\, d\omega$$

$$= \frac{1}{2\pi}\int_{-\pi}^{\pi} S_n(\omega)\, d\omega + \sum_{1}^{2p} \frac{\sigma_i^2}{2\pi}|H(e^{j\omega_i})|^2$$

$$= A + B(\phi)$$

where A is constant and equal to the noise power. So

$$\min_{\phi} J = A + \min_{\phi} B = A.$$

Therefore, the only minimizing condition is that the notches of the transfer function $H(e^{j\omega})$ be at frequencies $\omega_1, \omega_2, \cdots, \omega_{2p}$. This implies that minimization of J does indeed ensure an unbiased estimate of the sinusoidal frequencies. It is important to note that the analysis holds only for an ideal notch filter. The sharper the notch filter, the more accurate is the analysis, since $A(\phi)$ will tend to remain constant. The larger the bandwidth of the notch filter, the more $A(\phi)$, and, therefore, the estimate of ϕ, are susceptible to the shape of the colored noise spectrum.

In retrospect, the solution is not very surprising, as a notch filter appears to be a natural way to eliminate sinusoidal signals [10] and also is an unconstrained optimum filter for the problem. It may also be noted that the adaptive line enhancer implemented using an FIR structure reduces to such a band-rejection structure as the order gets larger [5]. Even the ARMALE reduces to a filter having such a response (see Section III).

For this problem, all the existing ALE's eventually converge to a band-rejection filter, the difference being the parameterization (ϕ). Different parameterizations have different advan-

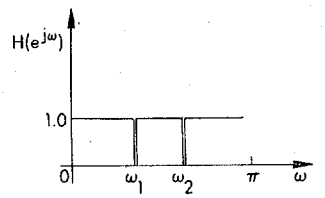

Fig. 2. Frequency response (magnitude) of an ideal notch filter.

tages and disadvantages, and the key may lie in choosing an appropriate parameterization. It appears natural that the parameterization should facilitate the formation of a band-rejection filter. Now to determine the parameterization, a time domain analysis is conducted.

III. Time Domain Modeling

We first examine the optimum Wiener filter for estimating n_t given y_t. Using the innovations approach, the filter for estimating n_t given y_t up to time t, $n_{t/t}$, can be shown to be as in Fig. 3. Section I is the whitening filter and Section II is the optimum filter based on the white noise observations [11]. Now a sinusoidal signal can be considered as the output of a system with no input, i.e., a self-generative process. As a sinusoid is a zero innovation process, the output of Section I v_t is due to n_t alone. Consequently, Section II is given by the canonical spectral factor $S_n^+(z)$ of $S_n(z)$, which contains the poles and zeros of $S_n(z)$ that are inside the unit circle.

Now, it is useful to split Section I into two parts as in Fig. 3. The first block whitens the noise and the second block removes the sinusoids. Now the output of block I, which is also the input to block II, consists of sinusoids in white noise. The sinusoids have the same frequencies as the input sinusoids except for a change in amplitude. As the ordering of the blocks is not important, block I cancels Section II, leaving block II as the only relevant section. So to derive the model it is adequate to determine block II, i.e., consider sinusoids in white noise. As long as the filter derived is independent of the amplitudes of the sinusoids, which it will be shown to be, the filter is good for the colored noise case. In this manner we can avoid the need for additional information regarding the noise. However, since there exists no rational innovations representation for sinusoidal signals in white noise, narrow-band signals in white noise are considered and a limiting approach is used to derive the approximate model.

A. Innovations Model for Narrow-Band Signals in White Noise

Assuming x_t is a sum of finite bandwidth signals, the autocorrelation $R_{yy}(k)$ for the signal y_t can be represented by [12]

$$R_{yy}(k) = \sigma^2 \delta(k) + \sum_1^p \sigma_i^2 \exp(-\gamma_i |k|) \cos(\omega_i k).$$

Here, σ^2 represents the variance of the white noise component n_t and σ_i^2, γ_i, ω_i represent the power, bandwidth, and frequency, respectively, of each of the p finite bandwidth signals in x_t.

For narrow-band signals where $\gamma_i \to 0$ and is much smaller than unity, the signal can be considered as the output of a linear system with a rational transfer function

$$H(z) = \frac{A \prod_1^{2p} (z - e^{-\mu_i + j\theta_i})}{\prod_1^{p} (z - e^{-\gamma_i + j\omega_i})(z - e^{-\gamma_i - j\omega_i})}$$

whose poles and zeros satisfy the following approximate relation [12].

$$\mu_i = (\gamma_i^2 + \mathrm{SNR}_i\, \gamma_i)^{1/2}$$

$$\mu_{i+p} = \mu_i \quad \text{for } i = 1, 2, \cdots, p$$

and

$$\theta_i = \omega_i \quad \text{and} \quad \theta_{i+p} = -\omega_i. \tag{7}$$

Here SNR_i stands for the signal-to-noise ratio for each component, i.e., $\mathrm{SNR}_i = \sigma_i^2/\sigma^2$. As the narrow-band signals approach sinusoidal signals, i.e., $\gamma_i \to 0$, the zeros and poles approach the unit circle. In the limit, the poles and zeros are on the unit circle at the same locations. However, for stability, the poles have to be kept away from the unit circle.

B. Notch Filter Parameterization

Combining the frequency domain analysis (6) and the time domain analysis (7), an approximate implementation of the notch filter in the time domain is the following: the poles and zeros are constrained to lie on the same radial line with the poles p_i lying in between the zeros z_i and the origin (Fig. 4), i.e., satisfying

$$p_i = \alpha z_i$$

$$= \alpha r_i e^{j\omega_i} \quad \text{for } i = 1, \cdots, 2p$$

where $0 \leqslant \alpha \leqslant 1$ and $0 \leqslant r_i \leqslant 1$. Normally $r_i \to 1$ for sinusoidal signals.

The notch filter transfer function is given by

$$H(z) = \frac{\prod_1^{2p} (1 - z_i z^{-1})}{\prod_1^{2p} (1 - \alpha z_i z^{-1})} \quad \text{where } 0 \leqslant \alpha < 1.$$

Because of the relationship between the zeros and the poles, it is not hard to see that the numerator and denominator coefficients of the filter are related. So the notch filter transfer function is given by

$$H(z) \triangleq \frac{W(z)}{W(\alpha z)}$$

where

$$W(z) = 1 - w_1 z^{-1} - w_2 z^{-2} - \cdots - w_{2p} z^{-2p} \tag{8}$$

$$W(\alpha z) = 1 - \alpha w_1 z^{-1} - \alpha^2 w_2 z^{-2} - \cdots - \alpha^{2p} w_{2p} z^{-2p} \tag{9}$$

and the weight vector $W = (w_1, w_2, \cdots, w_{2p})^T$. It can be seen that an ideal notch filter as considered in the last section is

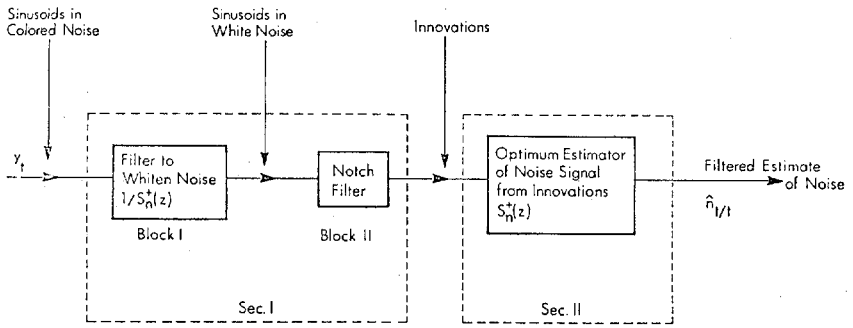

Fig. 3. Optimum filtered estimate of the noise signal.

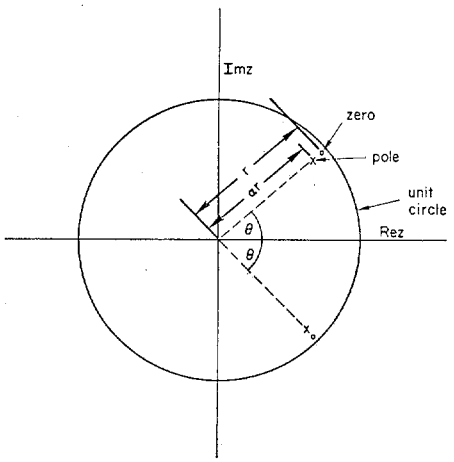

Fig. 4. Location of the pole zero pair for the notch filter.

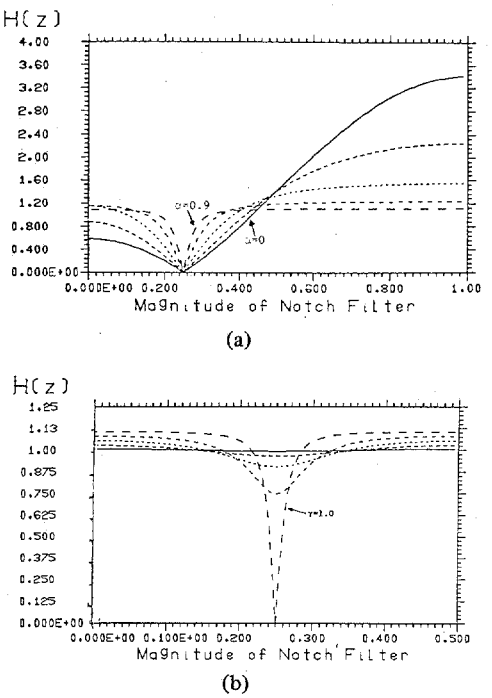

Fig. 5. (a) The magnitude of the notch filter response is shown as a function of the debiasing parameter α. Plot shows response for α equal to 0, 0.3, 0.6, 0.8, and 0.9 with the radius fixed at 1. (b) The magnitude of the notch filter response is shown as a function of the radius r. Plot shows response for r equal to 0.2, 0.4, 0.6, 0.8, and 1.0 with α fixed at 0.9.

better approximated as $\alpha \to 1$.[1] The closer α is to unity, the flatter the notch filter response (outside the notch frequency) will be, making the bias smaller. This is clearly evident from Fig. 5(a). So the parameter α is called the debiasing parameter. Also, the radius determines the nature of the notch filter as shown in Fig. 5(b).

Having determined the filter structure, we now turn our attention to the problem of estimating the filter coefficients. For this we note that the cost function is a nonlinear function of W. Furthermore, there exist efficient batch processing schemes (particularly for the white noise case) to determine the frequencies [13]-[15]. So it appears unlikely that employing conventional minimization schemes will result in a better frequency estimator. However, the employment of adaptive schemes can be useful for accurate frequency tracking. In addition, the filter can be easily configured to realize a line enhancer and obtain an enhanced version of the sinusoidal signals (see Fig. 6). The transfer function between input and $\hat{x}_t$, $1 - H(z)$, represents a tuned filter, as shown in Fig. 7 (for a second-order notch filter). The filter selects the sinusoidal signal and rejects the noise enhancing the signal-to-noise ratio. The enhancement clearly depends on the bandwidth (α) and the closeness of the center of the tuned filter to the input sinusoidal frequency. Hence, for the notch filter to be a good line enhancer, as $\alpha \to 1$, the center frequencies, roots of $W(z)$, should give good estimates of the input sinusoidal frequencies. Hence, we concentrate on the accuracy of the sinusoid frequency estimates obtained by the notch filter and use it as an indicator of the line enhancement properties. The rest of the paper investigates suitable adaptive schemes to estimate W, establishing their convergence properties and their closeness to the desired optimal W^*. Due to space limitations, more direct details on the enhancement and tracking properties are left to a later publication. As α is going to play a key role, we now examine its effect.

[1]For $\alpha = 1$, the notch filter given by (8) and (9) is defined, using a limiting approach, to be a function that is 1 almost everywhere as in (6).

IV. Relation to Existing Methods and the Role of α

From (8), the difference equation governing the notch filter is given by

$$\hat{n}_t - \alpha w_{t-1}\hat{n}_{t-1} - \cdots - \alpha^{2p} w_{2p}\hat{n}_{t-2p} = y_t - w_1 y_{t-1} - \cdots - w_{2p} y_{t-2p} \tag{10}$$

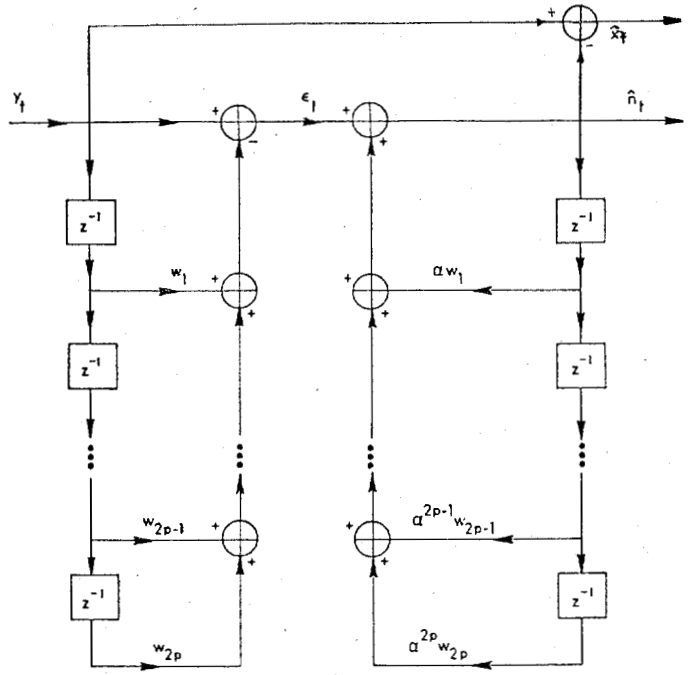

Fig. 6. The notch filter used as a line enhancer. The direct form II (zeros before the poles) is used to implement the transfer function.

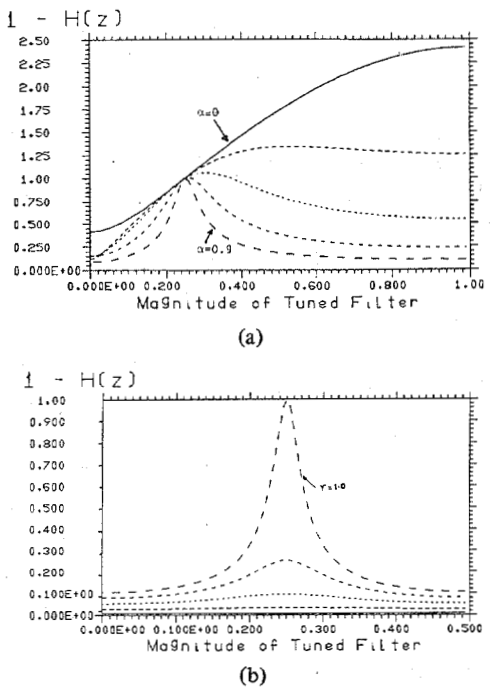

Fig. 7. (a) The magnitude of the tuned filter response is shown as a function of the debiasing parameter α. Plot shows response for α equal to 0, 0.3, 0.6, 0.8, and 0.9 with the radius fixed at 1. (b) The magnitude of the tuned filter response is shown as a function of the radius r. Plot shows response for r equal to 0.2, 0.4, 0.6, 0.8, and 1.0 with α fixed at 0.9.

i.e.,

$$W(\alpha z)\,\hat{n}_t = W(z)\,y_t, \qquad 0 \leqslant \alpha \leqslant 1$$

where $\hat{n}_t$, the filter output, represents in some sense an approximation to n_t. Rewriting (10) in compact form, we have

$$\hat{n}_t = y_t - W^T Y_t + W^T \Lambda \hat{N}_t \tag{11}$$

where vectors Y_t and $\hat{N}_t$ contain the past values and

$$\Lambda = \text{diag}\,(\alpha, \alpha^2, \cdots, \alpha^{2p}).$$

The variable α provides some flexibility in dealing with signals of finite but small bandwidth. Also, the special cases $\alpha = 0$ and $\alpha = 1$ are of special interest and are discussed next.

A. Relation to Existing Methods

For $\alpha = 0$ the transfer function, from (8), is equal to

$$H(z) = 1 - w_1 z^{-1} - w_2 z^{-2} - \cdots - w_p z^{-p}.$$

This is the popular linear prediction filter where w_i's represent the predictor coefficients or the coefficients of an autoregressive model. For $\alpha = 1$ we have, from (10), the difference equation governing the filter given by

$$\hat{n}_t = y_t - W^T Y_t + W^T \hat{N}_t.$$

This equation describes a special ARMA model from which Pisarenko's method can be derived [16]. Since the equation, by a limiting operation $\alpha \to 1$, represents a function that is 1 almost everywhere, Pisarenko's method can be interpreted as being an implementation of an extremely sharp notch filter.

B. Intermediate Values of α

Because of stability problems, the extreme value of α, $\alpha = 1$, is only of theoretical interest. For all practical situations, intermediate values of α will be used and its effect is now examined.

The weight vector is computed by minimizing the output power of the notch filter, $J[\alpha, W]$, which is a function of α and W. For every α, there exists a W or a set of W which minimizes J, thereby establishing a mapping from the set of α's to W's. It is this mapping that is studied to understand the behavior of the notch filter for intermediate values of α.

By using the implicit function theorem, it can be shown that there are neighborhoods around some values of α's for which the estimate W is continuous. However, the size of the open set is unknown. So, this existence result is only useful when local minima are known to exist. For example, it is useful when a filter of length smaller than required is used and an α close to 1 is chosen (as in the cascade and parallel implementations discussed later in Section VII). A stronger result can be obtained by restricting $J(\alpha, W)$ to have a single global minimum. With this restriction we have the following result.

Theorem: The mapping $f: \alpha \to W$ defined by $J(\alpha, f(\alpha)) = \inf_W J(\alpha, W)$ is continuous.

Proof: See Appendix A.

If there exists more than one minimum, then we have a set valued function for which an upper semicontinuous property can be proved. (Details in Appendix A.)

For the case of a single global minimum, since $W(z) = z^n - w_1 z^{n-1} - \cdots - w_n$ represents a continuous mapping of W onto the zeros in the z-plane, the following conclusion can be made using the above theorem. The optimum values of the filter coefficients give rise to zeros within the unit circle whose position is a continuous function of α: as α changes from 0 to 1, the zeros trace a continuous path. This is best understood by means of a simple numerical example. A signal consisting of a single sinusoid in white noise at an SNR of 0 dB is passed

through a second-order notch filter, and the weight vector W is obtained by numerically solving for the minimum of the steady-state cost function. The zeros of the polynomial $W(z)$ are plotted as a function of α in Fig. 8. The continuous movement of the zeros as a function of α seen in Fig. 8 is expected from the above theorem. This also indicates that there exists some flexibility in the choice of α. Now for our applications, line enhancement and tracking, we examine an adaptive scheme to estimate the weight vector W.

V. Adaptive Estimation

Before outlining the adaptation procedure, some preliminary rearrangement of the input-output description is useful. The usual input-output description, given by the difference equation (10), naturally leads to a zero pole filter implementation (direct form II), i.e., the zeros precede the poles. Rearranging the filter to have the poles before the zeros (direct form I), as in Fig. 9, results in the following input-output description:

$$\tilde{y}_t = y_t + \alpha w_1 \tilde{y}_{t-1} + \alpha^2 w_2 \tilde{y}_{t-2} + \cdots + \alpha^{2p} w_{2p} \tilde{y}_{t-2p} \quad (12)$$

or

$$\tilde{y}_t = y_t / W(\alpha z)$$

and

$$\hat{n}_t = \tilde{y}_t - w_1 \tilde{y}_{t-1} - w_2 \tilde{y}_{t-2} - \cdots - w_{2p} \tilde{y}_{t-2p}$$

or

$$\hat{n}_t = \tilde{y}_t - W^T \tilde{Y}_t$$

i.e.,

$$\hat{n}_t = W(z)\, \tilde{y}_t. \quad (13)$$

These equations will be used in the adaptive estimation schemes and subsequent manipulations.

A. Gauss-Newton Method

For the adaptation of the coefficients, the stochastic Gauss-Newton method is used [17]. The update is given by

$$W_t = W_{t-1} + \gamma_t R_t^{-1} \psi_t$$

$$R_t = R_{t-1} + \gamma_t [\psi_t \psi_t^T - R(t-1)].$$

Using an approximate Hessian matrix and the matrix inversion lemma [17], the following recursive update is obtained:

$$W_t = W_{t-1} + K_t \hat{n}_t$$

$$K_t = \frac{P_{t-1}\psi_t}{\lambda_t + \psi_t^T P_{t-1} \psi_t}$$

$$P_t = \left[P_{t-1} - \frac{P_{t-1}\psi_t \psi_t^T P_{t-1}}{\lambda_t + \psi_t^T P_{t-1}\psi_t} \right] \Big/ \lambda_t$$

$$\lambda_t = \lambda_0 \lambda_{t-1} + (1 - \lambda_0) \quad (14)$$

where ψ_t represents the gradient ($G1$ or $G2$).

$$G1: \psi_t = -\frac{dn_t}{dW} = \left(-\frac{\partial n_t}{\partial w_1} \cdots - \frac{\partial n_t}{\partial w_{2p}} \right)^T$$

$$= (\tilde{y}(t-1) - \alpha \tilde{n}(t-1), \cdots, \tilde{y}(t-2p) - \alpha^{2p} \tilde{n}(t-2p))^T$$

$$= (\tilde{Y}_t - \Lambda \tilde{N}_t) \quad (15)$$

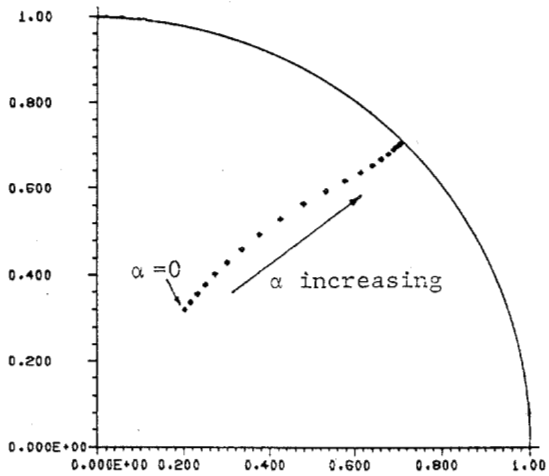

Fig. 8. The zeros of the notch filter are plotted as a function of the debiasing parameter α for values of α from 0 to 0.95 at an interval of 0.05. The SNR is 0 dB and the desired (optimum) position of root is on the unit circle at an angle of 45°.

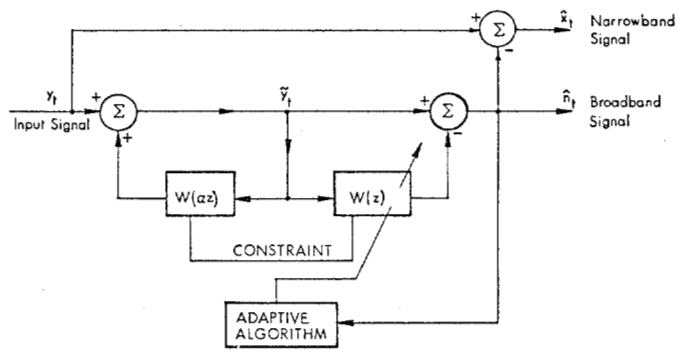

Fig. 9. The direct form I (poles before the zeros) implementation of the notch filter.

with

$$\Lambda = \text{diag}\,(\alpha, \alpha^2, \cdots, \alpha^{2p})$$

and μ is the fixed step size. $\tilde{y}_t$ is defined in (12) and $\tilde{n}_t$ is similarly defined, i.e.,

$$\tilde{n}_t = \hat{n}_t / W(\alpha z)$$

or

$$\tilde{n}_t = \hat{n}_t + \Lambda W^T \tilde{N}_t. \quad (16)$$

As can be seen, $\tilde{y}_t$ and $\tilde{n}_t$ are filter outputs and the derivatives can be obtained by simply using the delayed versions of $\tilde{y}_t$ and $\tilde{n}_t$. The gradient is labeled $G1$ for future reference, and now a simplification of $G1$ is proposed.

The issue of gradient truncation/simplification has been widely discussed in the literature [18], [19]. Truncated gradients often imply simpler computations and improved stability. However, they may not always converge to the optimal estimates [20], [21]. Here we present a simplified gradient which will be shown to be well behaved:

$$G2: \psi_t = \tilde{Y}_t.$$

This form of the gradient has been first used in [22] and is particularly suitable for the white noise case. An advantage of $G2$ is that it can be obtained from a direct form I implementation with no additional computation [see (12) and Fig. 9].

To ascertain that the parameter estimates obtained by using the adaptive algorithm (14) do converge to the desired optimum values, a convergence study is conducted. The analysis is conducted, following Ljung's work [23], by studying the associated ordinary differential equation. Here we summarize the main results. (The details are given in Appendix B.)

B. Gradient G1

This complete gradient has provable convergence properties. The stationary points of the algorithm satisfy

$$f(W^*) = 0.$$

This implies that for $W = W^*$

$$E[\hat{n}_t(\tilde{Y}_t - \Lambda\tilde{N}_t)] = 0.$$

Substituting for $\hat{n}_t$ from (13), we have

$$E[\hat{n}_t(\tilde{Y}_t - \Lambda\tilde{N}_t)] = E[-\tilde{Y}_t\tilde{Y}_t^T W_{t-1} + \tilde{y}_t\tilde{Y}_t - \Lambda\hat{n}_t\tilde{N}_t].$$

Again substituting for $\hat{n}_t$ from (16), we have

$$E[\hat{n}_t(\tilde{Y}_t - \Lambda\tilde{N}_t)] = E[-\tilde{Y}_t\tilde{Y}_t^T W_{t-1} + \tilde{y}_t\tilde{Y}_t + \Lambda\tilde{N}_t\tilde{N}_t^T\Lambda W_{t-1} - \tilde{n}_t\Lambda\tilde{N}_t].$$

The optimal estimate W^* is given by

$$W^* = R^{-1}(W^*)P(W^*)$$

where

$$R(W^*) = E[\tilde{Y}_t(W^*)\tilde{Y}_t^T(W^*)] - \Lambda E[\tilde{N}_t(W^*)\tilde{N}_t^T(W^*)]\Lambda = R_{\tilde{Y}} - \Lambda R_{\tilde{N}}\Lambda \quad (17)$$

and

$$P(W^*) = E[\tilde{y}_t(W^*)\tilde{Y}_t(W^*)] - \Lambda E[\tilde{n}_t(W^*)\tilde{N}_t(W^*)] = P_{\tilde{y}} - \Lambda P_{\tilde{n}}. \quad (18)$$

Here $\tilde{Y}_t(W^*)$ and $\tilde{N}_t(W^*)$ denote stationary processes obtained by the recursions generating the elements of $\tilde{Y}_t$ and $\tilde{N}_t$, respectively, for a W fixed at W^*. Since the noise and signal are independent,

$$R_{\tilde{Y}} = R_{\tilde{X}} + R_{\tilde{N}_a},$$

where $\tilde{x}_t = x_t/W(\alpha z)$ and $\tilde{n}_{at} = n_t/W(\alpha z)$. Note that n_t, not $\hat{n}_t$, is used in generation of $\tilde{n}_{at}$.

For $\alpha = 0$, this reduces to the familiar Wiener-Hopf equations. For $\alpha \simeq 1$, the filter is close to the optimum notch filter and W^* is expected to be close to the desired weight vector. Then an examination of $R(W^*)$ will show that it behaves like the covariance of the data from which the noise contribution has been removed. $R_{\tilde{N}}$ can be thought of as the noise component which is being subtracted from the measured covariance. This clean up effect is better as α increases. For α sufficiently close to 1, $R_{\tilde{N}}$ will closely approximate the noise component and

$$R(W^*) \simeq E[\tilde{X}_t\tilde{X}_t^T].$$

Similarly, $P(W^*) \simeq E[\tilde{x}_t\tilde{X}_t]$, where $\tilde{x}_t$ is the sinusoids amplified by their passage through $1/W(\alpha z)$. This amplification increases when the poles are close to the unit circle, i.e., as α increases, and can be viewed as a mechanism to enhance the signal-to-noise ratio. As $\alpha \to 1$, the solution obtained is the same as would be obtained by the linear prediction techniques for a signal where the sinusoids are amplified and the noise simultaneously reduced, as in (17) and (18). Consequently, the bias in the estimates reduces as α increases. These equations also indicate how, even for the colored noise case, the method is able to eliminate the effect of noise. The optimal solution W^* is also stable and the algorithm is globally asymptotically stable. These aspects are shown in Appendix B.

C. Gradient G2

This gradient, although computationally simpler, is more appropriate for the white noise case. The stationary points of the algorithm are again given by $f(W^*) = 0$. This implies for $W = W^*$

$$E[\hat{n}_t\tilde{Y}_t] = 0.$$

Substituting $\hat{n}_t$ from (13), we have

$$E[\tilde{Y}_t\tilde{Y}_t^T]\, W^* = E[\tilde{y}_t\tilde{Y}_t]$$

or

$$[R_{\tilde{X}} + R_{\tilde{N}_a}]\, W^* = P_{\tilde{x}} + P_{\tilde{n}_a} \quad (19)$$

where $R_{\tilde{N}_a}$ and $P_{\tilde{n}_a}$ are formed from the data $\tilde{n}_{at}$, where $\tilde{n}_{at}$ is as defined above. If n_t is white noise, then

$$R_{\tilde{N}_a}\Lambda W^* = P_{\tilde{n}_a} \quad (20)$$

and using (20), we have

$$R_{\tilde{X}}W^* = P_{\tilde{x}} - R_{\tilde{N}_a}(I - \Lambda)\, W^* \quad (21)$$

The last term for the white noise term is like a perturbation which goes to zero as $\alpha \to 1$. Even for finite α, $R_{\tilde{X}}$ and $P_{\tilde{x}}$ are from amplified sinusoids greatly reducing the bias. For the colored noise case, returning to (19), the situation is worse than the white noise case. But still for values of α close to one, the amplified sinusoids can greatly reduce the effect of the noise. From simulations, it has been noted that in a colored noise environment, there is usually a noticeable bias in the estimate, which decreases as α increases. The solution is also stable and almost[3] globally asymptotically stable. The success of the gradient depends on the fact that "$1 - W(z)/W(\alpha z)$" is positive real (almost 1) around the notch frequencies where the power due the sinusoids is concentrated. This structure in the problem makes the covergence proofs possible. (Details are given in Appendix B.)

VI. Simulation Results

Some simulation results are now presented to support the notch filtering method. A normalized sampling frequency of 2 Hz is assumed throughout, and SNR is defined for each sinusoid as $\text{SNR} \triangleq A^2/2R(0)$, where A is the amplitude of a sinusoid and $R(0)$ the noise power. Instead of the frequencies, the location of the zeros of $z^{2p}W(z)$ in terms of their radius and angle in the z-plane are tabulated. Because of the sampling frequency chosen, for a frequency of "f" Hz, the desired location of the z-plane zeros is on the unit circle (radius = 1) at an angle of $f\pi$ radians. First simulation results for the Gaussian white noise case are presented. For all simulations, a Gauss-Newton procedure [see (14)] is used with the following initial

[3] See Appendix B.

conditions: $P(0) = 0.01$, $\lambda_0 = 0.99$, and an initial value of 0.95 for λ_f.

Example 1: This example demonstrates the dependence on α, i.e., its debiasing effect. For this purpose, data consisting of a single sinusoid at a frequency of 0.25 Hz and an SNR of 3 dB is used. The results for a second-order notch filter after 25 trials for different data lengths are summarized in Table I. The estimates can be seen to improve continuously with α. This continuity property is expected from the theorem in Section III. For $\alpha = 0.95$, the larger variance is due to the nonlinearity of the gradient.

Example 2: Two sinusoids at frequencies of 0.25 and 0.70 Hz are used for this purpose. At an SNR of 3 dB and for $\alpha = 0.9$, the mean and standard deviation for 25 trials is summarized in Table II. As the frequencies are brought closer, the estimates deteriorate. Frequencies of 0.25 and 0.30 Hz at an SNR of 3 dB are used. The results of 25 trials for a record of 256 samples and an α equal to 0.9 indicate that there is a tendency to identify one sinusoid. As the data length increases, the frequencies are correctly identified. Table III, where the results of 25 trials for a data record of 512 samples are summarized, illustrates this behavior.

Next, some colored noise simulations are conducted. For this purpose the sinusoids are corrupted by an AR signal.

Example 3: The AR noise signal has a power spectrum given by

$$\frac{\sigma^2}{|1 - 0.309\, e^{-j\omega} + 0.25\, e^{-2j\omega}|^2}.$$

Three sinusoids, at frequencies of 0.25, 0.70, and 0.8 Hz, of equal amplitude at an SNR of 3 dB are used. The actual spectrum is shown in Fig. 10, and the results for a data length of 512 samples averaged over 25 trials are summarized in Table IV. These simulations clearly indicate the potential of the notch filtering approach.

VII. Effect of Filter Length

In the above simulations and analysis, it is assumed that the size of the filter is appropriately chosen. Here we consider the effect of having either a larger or smaller filter.

A. Effect of Large Filter Length

In the event the filter length exceeds the required length, a frequency domain analysis can be done to determine its effect. We have seen that the output power of the notch filter is given by [see (4)]: $J = A(\phi) + B(\phi)$. The signal contribution $B(\phi)$ can be made zero if an ideal notch filter ($\alpha = 1$) is used. In practice, if α is sufficiently close to 1 and if the order of the filter is $2L > 2p$, p being the number of sinusoids, then $2p$ zeros of the notch filter will be fixed by the sinusoids. The rest of the zeros/coefficients will adjust to minimize the noise contribution $A(\phi)$, which for the white noise case is given by

$$A(\phi) = \sigma^2 \oint H(z)\, H^*(z^{-1})\, \frac{\Pi(z - z_i)\, \Pi(1 - z z_i^*)\, dz}{\Pi(z - \alpha z_i)\, \Pi(1 - \alpha z z_i^*)\, z}$$

where $H(z)$ is fixed by the p sinusoids. Assuming that the value of α is chosen sufficiently close to 1, the interaction between pole zero pairs can be ignored. Then the contribution to $A(\phi)$ from the zeros $z_i = 2p + 1, \cdots, 2L$ that have yet to be determined is given by

$$C = -\sum_{2p+1}^{L} \sigma^2 (1 - \alpha)/\alpha((1 - \alpha|z_i|^2)/(1 - \alpha^2|z_i|^2)).$$

C is minimum when the z_i's are zero, thereby implying that the excess zeros go to the origin. In practice, they have no effect once they are sufficiently inside the unit circle [see Fig. 5(a) and (b)]. As a result, the excess pole zero pairs tend to go inside the unit circle till they become ineffective. To study this effect, a fourth- and sixth-order filter are used on the data of Example 1. 25 trials were performed with $\alpha = 0.9$ and in all these cases, only one zero of $z^{2p} W(z)$ had a radius close to unity and the rest less than 0.8. The excess zeros tend to go further away from the unit circle as more data are processed.

B. Local Minima

Next, considering the case where we have a filter of order lower than required, some interesting facts are observed. Without any loss of generality, let us assume that a second-order filter is used. Then, using a frequency domain analysis as before (see Section II), it can be shown that there exist local minima which correspond to the removal of one sinusoid. In the practical case, by the implicit function theorem, it holds only for values of α close to 1. So generalizing, a notch filter of order $2r$ will remove r of the p sinusoids for $r \leqslant p$. This unique behavior of the notch filter lends itself to the following interesting implementations.

C. Implementation of Cascade Forms

One implementation is cascade implementation, shown in Fig. 11. When α is sufficiently close to 1, the bandwidths of each section have negligible or no overlap. By the local minima property, each section can eliminate one sinusoid. So, instead of using the overall output to adjust the fiber coefficients, this frequency decoupling property enables each section to be independently minimized. The first section cancels one sinusoid and transmits the rest of the signal, with little or no distortion, to the second section, which cancels a second sinusoid if present. The output of Section I is used to adaptively adjust its coefficients, while the output of Section II is used to determine the coefficients of the second section. This property is unique to the notch filter and similiar attempts in other contexts have not been so successful [24].

D. Parallel Form Implementation to the Notch Filter

Another useful implementation is the parallel form implementation. It is not a parallel form in the usual sense. Here, a number of second-order notch filters are operated in parallel, as shown in Fig. 12. The local minima property, discussed before, guarantees that each section removes one sinusoid. Furthermore, the local minima which each section converges to depends on the starting point. Initialization of each second-order section is straightforward as only two parameters are involved. Therefore, the entire frequency domain can be divided into small zones with each zone being monitored by a second-order notch filter. The exact number of bins and their size are dependent on the problem at hand.

The cascade and parallel forms discussed above have been successfully verified through simulations and offer several

TABLE I
SIMULATION RESULTS FOR ONE SINUSOID IN WHITE NOISE AS A FUNCTION OF α AND FOR A DATA LENGTH OF 256 AND 512

value of α	$N = 256$				$N = 512$			
	Radius		Angle (rad)		Radius		Angle (rad)	
	mean	standard deviation	mean	standard deviation	mean	standard deviation	mean	standard deviation
0	0.55812	0.05078	0.30911 π	0.00745 π	0.53651	0.05165	0.30882 π	0.00492 π
0.1	0.61563	0.05065	0.30074 π	0.00715 π	0.59664	0.05019	0.30105 π	0.00492 π
0.2	0.68584	0.04886	0.29031 π	0.00701 π	0.67121	0.04602	0.29074 π	0.00499 π
0.3	0.76456	0.04282	0.27813 π	0.00628 π	0.75598	0.03735	0.27810 π	0.00437 π
0.4	0.84003	0.03189	0.26637 π	0.00463 π	0.83650	0.02546	0.26574 π	0.00301 π
0.5	0.89949	0.02049	0.25783 π	0.00299 π	0.89816	0.01507	0.25693 π	0.00177 π
0.6	0.94064	0.01222	0.25320 π	0.00212 π	0.94003	0.00840	0.25232 π	0.00105 π
0.7	0.96843	0.00700	0.25128 π	0.00172 π	0.96824	0.00445	0.25056 π	0.00059 π
0.8	0.98636	0.00425	0.25059 π	0.00133 π	0.98670	0.00205	0.25011 π	0.00030 π
0.9	0.99405	0.00484	0.25033 π	0.00108 π	0.99635	0.00106	0.25006 π	0.00019 π
0.95	0.96965	0.03157	0.25064 π	0.02064 π	0.97502	0.02830	0.25007 π	0.01455 π

The true values are: radius = 1 and angle = 0.25 π rad.

TABLE II
SIMULATION RESULTS FOR TWO FAIRLY SEPARATE SINUSOIDS IN WHITE NOISE FOR α FIXED AT 0.9 AND FOR DATA LENGTHS OF 256 AND 512

		$N = 256$		$N = 512$	
		mean	standard deviation	mean	standard deviation
freq. 1	radius	0.99649	0.00218	0.99689	0.00068
	angle	0.25017 π	0.00078 π	0.25003 π	0.00015 π
freq. 2	radius	0.99669	0.00149	0.99710	0.00066
	angle	0.69995 π	0.00055 π	0.69997 π	0.00021 π

The true values: radii = 1, angles = 0.25 π and 0.70 π.

TABLE III
SIMULATION RESULTS FOR TWO CLOSELY SPACED SINUSOIDS IN WHITE NOISE FOR α FIXED AT 0.9 AND FOR A DATA LENGTH OF 512

	Radius		Angle	
	mean	standard deviation	mean	standard deviation
freq. 1	0.99011	0.01636	0.24973 π	0.00140 π
freq. 2	0.99094	0.00946	0.30070 π	0.00108 π

The true values: radii = 1, angles = 0.25 π and 0.30 π.

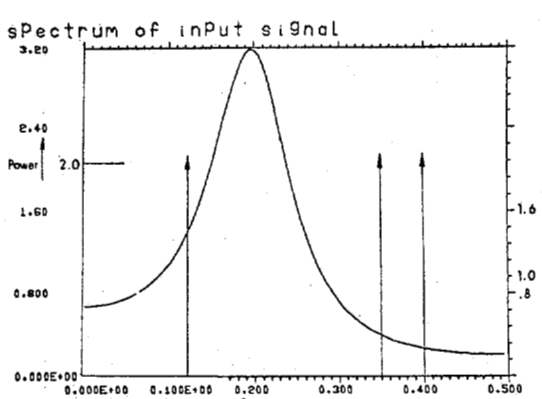

Fig. 10. The power spectrum of the signal used in example 3.

TABLE IV
SIMULATION RESULTS OF THREE SINUSOIDS IN AR NOISE FOR AN $\alpha = 0.9$ AND DATA LENGTH OF 512

	Radius		Angle	
	mean	standard deviation	mean	standard deviation
freq. 1	0.99379	0.00167	0.25047 π	0.00040 π
freq. 2	0.99615	0.00382	0.69970 π	0.00098 π
freq. 3	0.99744	0.00191	0.80000 π	0.00035 π

The true values are: radii = 1, angles = 0.25 π, 0.70 π, 0.80 π.

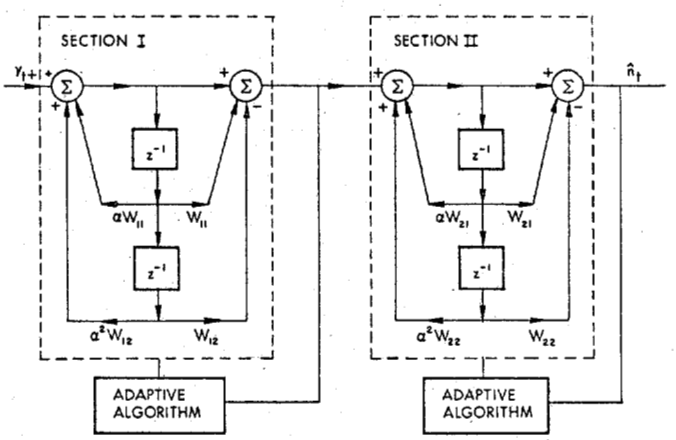

Fig. 11. Cascade implementation of an adaptive notch filter.

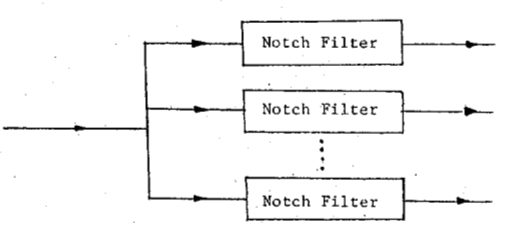

Fig. 12. Parallel implementation of the notch filter.

advantages. First, all roots are sought simultaneously, alleviating the problem of polynomial deflation [24]. Second, the partitioning of the whole problem into smaller problems and their decoupled nature provides an effective way to use multiprocessors for real-time processing. With VLSI, this alternative appears to be very promising. Lastly, and perhaps most importantly, *a priori* information and constraints on the location of the roots can be easily incorporated. A simple second-order filter can be used to track a single sinusoid, ignoring the rest of the spectrum.

VIII. Conclusions

In this paper, we have developed a new filter which is intuitively appealing and has a number of interesting properties. The filter developed requires a much smaller filter length as opposed to the popular FIR-ALE. Admittedly, the computational complexity of the update is larger, but with the small filter length, the advent of VLSI, and cheaper hardware, this may be less of an issue. The algorithm requires some vector-outer products which are amenable to parallel and real-time processing [27]. In addition, the truncated gradient and a direct form I implementation can reduce the complexity further.

Compared to the ARMA model, the new filter requires half the number of parameters. Also, it requires no additional order to handle a more general noise (colored) environment. The reduced number of parameters also result in reduced adaptation error. In addition, the notch filter enables simplification of the gradient and has an interesting decoupling property. In the presence of prior information, unlike the FIR-ALE and the ARMALE, the notch filter can be used to lock on to one frequency, ignoring the rest of the spectrum. The use of a simple second-order section greatly reduces the computational complexity and has better numerical behavior.

In conclusion, there is enough evidence to believe that the notch filter promises to be a viable candidate for the problem of line enhancement and tracking. Our results indicate some of the potential.

Appendix A

Dependence on α

Here we present results that explain the behavior of the notch filter for intermediate values of α. The results are presented in a general form first, and interpreted later for the problem at hand.

Let S be a subset of R^n such that $W \in S$, if the roots z_i of the polynomial

$$z^n - w_1 z^{-1} - \cdots - w_n = \Pi(z - z_i)$$

are inside or on the unit circle, i.e., $|z_i| \leq 1$. This set S is compact.

Lemma 1: If g is a real valued function which is continuous on the Cartesian product of $X \times Y$ of two Euclidean spaces, and if $\bar{Y}$ is a bounded subset of Y, then $\inf_{y \in Y} g(x, y)$ is a continuous function of x.

The proof is a minor modification of results obtained in [26] and given in [27].

Assuming that g has a unique infimum in $\bar{Y}$ for every x, we have the following result.

Theorem 1: The function $f: X \to \bar{Y}$ defined by $g(x, f(x)) = \inf_{y \in Y} g(x, y)$, is a continuous function.

Proof: Consider a sequence $\{x_n\}$, $x_n \to x_0$ in X. From Lemma 1, $g(x, f(x))$ is a continuous function of x. So,

$$g(x_n, f(x_n)) \to g(x_0, f(x_0)).$$

But $g(x, y)$ is a continuous function of x and y. So,

$$\lim_{n \to \infty} g(x_n, f(x_n)) = g(\lim_{n \to \infty} x_n, \lim_{n \to \infty} f(x_n))$$

$$= g(x_0, \lim_{n \to \infty} f(x_n))$$

$$g(x_0, f(x_0)) = g(x_0, \lim_{n \to \infty} f(x_n)).$$

By assumption there is a unique infimum. So,

$$f(x_0) = \lim_{n \to \infty} f(x_n)$$

i.e., f is continuous.

If there exists more than one infimum, we have a set valued function for which an upper semicontinuous property exists [27].

A. Interpretation

The output power is minimized to obtain the coefficients of the notch filter. The output power $J(\alpha, W)$ is a function of $W \in S$ and $\alpha \in [0, 1]$. It is similar to the function $g(x, y)$ described above with α playing the same role as x and W that of y. Based on the above results the following conclusions can be drawn.

The $\inf J(\alpha, W)$ is a continuous function of α (Lemma 1). Second, the function $f: \alpha \to S$ such that $J(\alpha, f(\alpha)) = \inf_W J(\alpha, W)$ is continuous if there exists a unique infimum (Theorem 1). The significance of this continuity has been explained in Section IV through a numerical example (cf. Fig. 8).

Appendix B

Convergence Analysis

For this analysis it is assumed that the order of the filter matches the actual model order and that the step size γ_t decreases and eventually becomes zero. Following the convergence results in [23], the update given by (14) can be associated with the following ordinary differential equation:

$$\frac{dW(\tau)}{d\tau} = R^{-1}(\tau) f(W(\tau))$$

$$\frac{dR(\tau)}{d\tau} = G(W(\tau)) - R(\tau) \tag{B.1}$$

where

$$f(W) = E\{\psi(t; W)\, \hat{n}(t; W)\}$$

$$G(W) = E\{\psi(t; W)\, \psi(t; W)^T\}.$$

Here $\hat{n}(t; W)$ and $\psi(t; W)$ denote the stationary processes obtained by the recursions generating $\hat{n}(t)$ and $\psi(t)$, respectively, for a fixed W.

As proved in [17], [23]:

- Only stable stationary points of (B.1) are possible convergence points of the algorithm. If $W(t) \to W^*$ with nonzero probability, then

$$f(W^*) = 0$$

and the matrix

$$G^{-1}(W^*)\left.\frac{df(W)}{dW}\right|_{W=W^*}$$

has all eigenvalues in the left half plane.

- Global asymptotical stability of a stationary solution $(W^*, G(W^*))$ of (B.1) implies that $W(t) \to W^*$ *w.p.*1. as $t \to \infty$ for the algorithm.
- The trajectories of the differential equation can be interpreted as the asymptotic paths of the algorithms.

For sinusoids in white noise we note that as $\alpha \to 1$, the notch filter describes the inverse filter more and more accurately. For this analysis, we assume α is sufficiently close to 1. If we further assume that the difference equation (10) provides an exact description of the inverse filter for the model generating the process, the convergence of $G1$ is easily established following the procedure used for the recursive maximum likelihood approach [17]. Here, we assume that the notch filter is an approximation and examine the consequences of the approximation. For the stability of the stationary points, we need to examine

$$\frac{df}{dW} = E\frac{d\psi}{dW}\hat{n}(t; W) + E\psi(t; W)\left(\frac{d\hat{n}(t; W)}{dW}\right)^T. \tag{B.2}$$

If the notch filter was an exact description, at $W = W^*$, the first term on the right-hand side is equal to 0 as $\hat{n}(t; W^*)$ is independent of the past. For the approximate case, we consider the second term and show that it is likely to dominate.

$$E\psi(t; W^*)\left(\frac{d\hat{n}(t; W)}{dW}\right)^T\Bigg|_{W^*} = E[\psi(t; W^*)\,\psi(t; W^*)^T] = -G(W^*).$$

Using the arguments used before, $G(W^*)$ represents the covariance of the amplified sinusoids and is proportional to $1/(1-\alpha)^2$. As α increases, the model gets accurate. Consequently, the first term on the right-hand side of (B.2) goes to 0 while the second term increases as $1/(1-\alpha)^2$ and is likely to dominate. Then we have

$$G^{-1}(W^*)\left.\frac{df}{dW}\right|_{W^*} \simeq G^{-1}(W^*)\,G(W^*) = -1,$$

indicating that the solution is stable. Now $V(W, \boldsymbol{R}) = \frac{1}{2}E(\hat{n}(t; W)^2)$ can be used as a Lyapunov function to establish the global stability of the differential equation.

The convergence of gradient $G2$ is along the same lines and follows closely the ideas in [28]. Again neglecting the first term, we have

$$\begin{aligned}\tilde{G}(W^*) &\simeq \left.\frac{df}{dW}\right|_{W^*}\\ &= E(\tilde{Y}_t(\tilde{Y}_t - \Lambda\tilde{N}_t)^T)\\ &= E(\tilde{Y}_t(\Lambda\tilde{Y}_t - \Lambda\tilde{N}_t)^T) + E(\tilde{Y}_t\tilde{Y}_t^T)(I-\Lambda)\\ &= E(\tilde{Y}_t(\tilde{Y}_t - \tilde{N}_t)^T)\Lambda + E(\tilde{Y}_t\tilde{Y}_t^T)(I-\Lambda)\\ &= E(\tilde{Y}_t((1 - W(z)/W(\alpha z))\,\tilde{Y}_t)^T)\Lambda\\ &\quad + E(\tilde{Y}_t\tilde{Y}_t^T)(I-\Lambda)\\ &= G_1\Lambda + G_2\end{aligned}$$

where

$$G_1 = E(\tilde{Y}_t((1 - W(z)/W(\alpha z))\,\tilde{Y}_t)^T)$$

and

$$G_2 = E(\tilde{Y}_t\tilde{Y}_t^T)(I-\Lambda).$$

To prove stability we need to show that

$$G^{-1}(W^*)\left.\frac{df}{dW}\right|_{W^*} = -G^{-1}(W^*)\,\tilde{G}(W^*)$$

has its eigenvalues in the left half plane, for which we make repeated use of the following result [28]: if S is positive semidefinite and $G + G^T$ is positive semidefinite, then SG has all its eigenvalues in the right half plane (including the imaginary axis).

Now, for stability it is sufficient to show that $(\tilde{G}(W^*) + \tilde{G}^T(W^*))$ is positive semidefinite. It is easy to see that $G_2 + G_2^T$ is positive semidefinite (this can be proved by using the above result from [30]). However, we still have to prove the positive semidefiniteness of $G_1 + G_1^T$. Then, positive semidefiniteness of $G_1\,\Lambda$ is again established using the above result. To check for the positive definiteness of G_1, we consider the quadratic form

$$L^T(G_1 + G_1^T)L = 2E(z_t\tilde{z}_t)$$

where

$$z_t = L^T\tilde{Y}_t \quad \text{and} \quad \tilde{z}_t = z_t(1 - W(z)/W(\alpha z))$$

are scalars. Now the covariance $E(z_t\tilde{z}_t)$ is given by

$$E(z_t\tilde{z}_t) = \frac{1}{2\pi}\int(1 - W(z)/W(\alpha z))\,\phi_z(\omega)\,d\omega$$

where $\phi_z(\omega)$ is the spectral density of the stationary process z_t. Since $\phi_z(\omega)$ and $E(z_t\tilde{z}_t)$ are real, we have

$$E(z_t\tilde{z}_t) = \frac{1}{2\pi}\int\mathrm{Re}\,(1 - W(z)/W(\alpha z))\,\phi_z(\omega)\,d\omega.$$

So $E(z_t\tilde{z}_t)$ is positive if $1 - W(z)/W(\alpha z)$ is positive real. However, in this case $1 - W(z)/W(\alpha z)$ need be positive only at the discrete frequencies at which $\phi_z(\omega)$ has its power concentrated, due to the amplified sinusoidal signals present in $\tilde{Y}_t$. This is true since, at the sinusoidal frequencies, $W(z)/W(\alpha z)$ has its notches and is close to 0. So Re $(1 - W(z)/W(\alpha z))$ is positive, and almost equal to 1 at the sinusoid frequencies, and nearly 0 at the other frequencies. Even if there is a negative contribution from the noise power at the other frequencies, it will be compensated for by the positive contribution at and near the sinusoid frequencies. Only at low SNR could there be a problem which can be compensated for by picking a larger value of α. So, $G_1 + G_1^T$ is positive semidefinite, ensuring positive semidefiniteness of $(\tilde{G}(W^*) + \tilde{G}^T(W^*))$ and, hence, the stability of the stationary point.

By examining the differential equation, linearized about the optimal solution, local convergence can be established [28]. $\tilde{G}(W)$ is again examined to establish global convergence properties.

$$\tilde{G}(W) = G_1\Lambda + G_2.$$

Here we need to show that G_1 is positive definite for all W. The case $W = W^*$ has been shown above. When W is far from

W^*, and inside the unit circle, then $W(z)/W(\alpha z)$ is almost 1 for values of α close to 1 [see Fig. 5(a)]. So $1 - W(z)/W(\alpha z)$ is nearly 0 and the presence of G_2 ensures positive definiteness. Although the arguments are not rigorous, it appears likely that for values of α close to 1, global convergence properties exist or, at least, $G2$ has a large domain of convergence. From the simulations conducted, the adaptive procedure has always converged to the optimal solution.

References

[1] B. Widrow *et al.*, "Adaptive noise cancelling: Principles and applications," *Proc. IEEE*, vol. 63, pp. 1692–1716, Dec. 1975.

[2] L. J. Griffiths, "Rapid measurement of digital instantaneous frequency," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-23, pp. 207–222, Apr. 1975.

[3] J. R. Treichler, "Transient and convergent behavior of the adaptive line enhancer," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-27, pp. 53–62, Feb. 1979.

[4] —, "The spectral line enhancer–The concept, an implementation and an application," Ph.D. dissertation, Stanford Univ., Stanford, CA.

[5] J. L. Zeidler, E. H. Satorius, D. M. Chabries, and H. T. Wexler, "Adaptive enhancement of multiple sinusoids in uncorrelated noise," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-26, pp. 240–254, June 1978.

[6] J. T. Rickard and J. Zeidler, "Second order output statistics of the adaptive line enhancer," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-27, pp. 31–39, Feb. 1979.

[7] N. J. Bershad *et al.*, "Tracking characteristics of the LMS adaptive line enhancer–Response to a linear chirp signal in noise," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-28, pp. 506–516, Oct. 1980.

[8] S. M. Kay, "The effect of noise on the autoregressive spectral estimator," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-27, pp. 478–485, Oct. 1979.

[9] B. Friedlander, "A recursive maximum likelihood algorithm for ARMA line enhancement," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-30, pp. 651–657, Aug. 1982.

[10] J. R. Glover, "Adaptive noise cancelling applied to sinusoidal interferences," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-25, pp. 484–491, Dec. 1977.

[11] T. Kailath, *Lectures in Linear Least-Squares Estimation*. New York: Springer-Verlag, 1976.

[12] S. T. Alexander, E. H. Satorius, and J. R. Zeidler, "Linear prediction and maximum entropy spectral analysis of finite bandwidth signals in noise," in *Proc. Int. Conf. Acoust., Speech, Signal Processing*, 1978, pp. 188–191.

[13] D. W. Tufts and R. Kumaresan, "Estimation of frequencies of multiple sinusoids: Making linear prediction perform like maximum likelihood," *Proc. IEEE*, vol. 70, pp. 975–989, Sept. 1982.

[14] J. A. Cadzow, "Spectral estimation: An overdetermined rational model equation approach," *Proc. IEEE*, vol. 70, pp. 907–939, Sept. 1982.

[15] S. Y. Kung, K. S. Arun, and D. V. Bhaskar Rao, "State-space and singular value decomposition-based approximation methods for the harmonic retrieval problem," *J. Opt. Soc.*, vol. 73, pp. 1799–1811, Dec. 1983.

[16] V. F. Pisarenko, "The retrieval of harmonics from a covariance function," *Geophys. J. Roy. Astron. Soc.*, vol. 33, pp. 347–366, 1973.

[17] T. Soderstrom, L. Ljung, and I. Gustavsson, "A theoretical analysis of recursive identification methods," *Automatica*, vol. 14, pp. 231–244, 1978.

[18] V. Panuska, "A stochastic approximation method for on-line identification on linear systems using adaptive filtering," in *Proc. Joint Automat. Contr. Conf.*, 1968.

[19] P. L. Feintuch, "An adaptive recursive LMS filter," *Proc. IEEE*, vol. 64, pp. 1622–1624, Nov. 1976.

[20] L. Ljung, T. Soderstrom, and I. Gustavsson, "Counterexamples to general convergence of a commonly used recursive method," *IEEE Trans. Automat. Contr.*, vol. AC-20, pp. 643–652, Oct. 1975.

[21] C. R. Johnson and M. G. Larimore, "Comments on and additions to 'An adaptive recursive LMS filter,' " *Proc. IEEE*, pp. 1399–1401, Sept. 1977.

[22] P. A. Thompson, "A constrained recursive filter for enhancing narrow-band signals in white noise," Ph.D. dissertation, Stanford Univ., Stanford, CA, Feb. 1979.

[23] L. Ljung, "Analysis of recursive stochastic algorithms," *IEEE Trans. Automat. Contr.*, vol. AC-22, pp. 551–575, Aug. 1977.

[24] L. B. Jackson and S. L. Wood, "Linear prediction in cascade form," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-26, pp. 518–528, Dec. 1978.

[25] S. Y. Kung, K. S. Arun, R. J. Gal-Ezer, and D. V. Bhaskar Rao, "Wavefront array processor: Language, architecture, and applications," *IEEE Trans. Comput.*, vol. C-31, pp. 1054–1066, Nov. 1982.

[26] E. L. Jennrich, "Asymptotic properties of non-linear least square estimators," *Ann. Math. Statist.*, vol. 40, pp. 633–643, 1969.

[27] D. V. Bhaskar Rao and S. Y. Kung, "Parameter estimation aspects and the role of bandwidth in a notch filter," in *Proc. 21st CDC*, Orlando, FL, Dec. 1982, pp. 1305–1307.

[28] L. Ljung, "On positive real transfer functions and the convergence of some recursive schemes," *IEEE Trans. Automat. Contr.*, vol. AC-22, pp. 539–551, Aug. 1977.

D. V. Bhaskar Rao received the B.Tech. degree in electronics and electrical communication engineering from the Indian Institute of Technology, Kharagpur, India, in 1979, and the M.S. and Ph.D. degrees in electrical engineering from the University of Southern California, Los Angeles, in 1981 and 1983, respectively.

From 1979 to 1983 he was a Teaching and Research Assistant in the Department of Electrical Engineering–Systems at the University of Southern California. Since July 1983 he has been with the Department of Applied Mechanics and Engineering Sciences at the University of California, San Diego, as an Assistant Professor. His research interests are in the areas of digital signal processing, estimation theory, and VLSI.

Sun-Yuan Kung (M'77–SM'84) received the B.S. degree in 1971 from the National Taiwan University, Taipei, the M.S. degree in 1974 from the University of Rochester, Rochester, NY, and the Ph.D. degree in 1977 from Stanford University, Stanford, CA, all in electrical engineering.

In 1974 he joined the Amdahl Corporation, Sunnyvale, CA, as an Associate Engineer in LSI design and simulation. From 1974 to 1977, he was a Research Assistant at the Information Systems Laboratories, Stanford University. In July 1977, he joined the Department of Electrical Engineering–Systems at the University of Southern California, Los Angeles, where he is currently an Associate Professor. In 1984, he was a Visiting Professor at Stanford University and, in the same year, a Visiting Professor at the Delft University of Technology, Delft, The Netherlands. His research interests are in the areas of approximation theory in linear systems, digital signal processing, modern and high-resolution spectrum analysis, parallel array processors, and VLSI supercomputing for signal processing. Since 1981, he has been in charge of the ONR Selected Research Project on the development of massively parallel signal processors. He is an editor of an advanced research book entitled *VLSI and Modern Signal Processing* (Englewood Cliffs, NJ: Prentice-Hall, 1984, to be published).

Dr. Kung served as the General Chairman of the USC Workshop on VLSI Signal Processing, Los Angeles, in November 1982. Currently he serves on the IEEE ASSP Technical Committee on VLSI, and is the Associate Editor for the VLSI area in the IEEE TRANSACTIONS ON ACOUSTICS, SPEECH, AND SIGNAL PROCESSING. He is a member of the Association for Computing Machinery.