

A Minimal Parameter Adaptive Notch Filter With Constrained Poles and Zeros

ARYE NEHORAI, MEMBER, IEEE

Abstract—A new algorithm is presented for adaptive notch filtering and parametric spectral estimation of multiple narrow-band or sine wave signals in an additive broad-band process. The algorithm is of recursive prediction error (RPE) form and uses a special constrained model of infinite impulse response (IIR) with a minimal number of parameters. The convergent filter is characterized by highly narrow bandwidth and uniform notches of desired shape. For sufficiently large data sets, the variances of the sine wave frequency estimates are of the same order of magnitude as the Cramer-Rao bound. Results from simulations illustrate the performance of the algorithm under a wide range of conditions.

I. INTRODUCTION

IN certain applications of signal processing, it is desired to eliminate narrow-band or sine wave components for observed time series. Examples are in the areas of radar, communications, control, biomedical engineering, and others. A practical situation is the case in which a sinusoidal power grid pickup of 50 or 60 Hz corrupts a measurement signal. Usually, this task is accomplished by fixed notch filters characterized ideally by a unit gain at all frequencies, except at the sine wave frequencies where their gain is zero. However, when the sine wave frequencies are unknown or possibly time varying, the fixed filters are not applicable, and it is necessary to apply adaptive notch filters whose characteristics are determined by the observed time series.

In this paper we shall design a new adaptive notch filter of infinite impulse response (IIR) for elimination of multiple unknown sine waves imbedded in a broad-band signal. The desirable pole-zero configuration of such a filter is schematically illustrated in Fig. 1 for the special case of a single sine wave in an additive broad-band signal (only the positive imaginary part is shown for convenience). As illustrated in the figure, the zeros of the filter have to lie on the unit circle at the sine wave frequency, and the poles have to be inside the unit circle at the same angles and as close as possible to the zeros. This description is similarly extended to multiple sine waves.

Previous adaptive IIR notch filters were designed by Thompson [1], Kung and Rao [2], [3], and Friedlander and Smith [4]. These filters usually satisfied only part of the desired properties of the notch filter and required $2n$ parameters, where n is the number of input sine waves.

For instance, the IIR filters of [1] and [3] did not constrain the zeros to be on the unit circle and required $2n$ parameters. Since the zeros were not constrained to be on the unit circle, their convergent angles were, in general, slightly different from the sine wave frequencies, and their notch bandwidths were larger than in the case of zeros on the unit circle. Another method in [2] and [3] was based on estimating the second-order factors of the IIR transfer function. This approach suffers from high nonlinearity in the minimization problem, which complicates the algorithm and deteriorates its performance, particularly as the number of sine waves increases. Friedlander and Smith [4] designed an adaptive notch filter (ANF) with zeros on the unit circle and soft-constrained poles. This algorithm required $2n$ parameters and a surplus pole-zero pair or notch at the Nyquist frequency. The use of soft-constrained poles may cause numerical problems due to near pole-zero cancellations on the unit circle, as well as different notch bandwidths than required.

The proposed ANF of this paper will satisfy the requirements of the notch filter, and will use a minimal number of parameters equal to n , the number of the input sine waves or narrow-band signal components. Other advantages of this algorithm over previous schemes include its computational efficiency, stability, more accurate results, numerical robustness, faster convergence, and better controllability on the performance. Our constraints on both poles and zeros will render sharper notches than before possible. Both zeros and poles will converge to the sine wave frequencies and their desired moduli, which was not the case in previous papers. We shall introduce a special time-varying pole modulus that will increase the robustness of the algorithm and improve its convergence rate. A more detailed account for the advantages of the proposed algorithm will be given in the next section. These advantages will be illustrated in computer simulation results. The following algorithm is also applicable to adaptive line enhancement, as will be explained later.

In Section II we consider the special structure of the proposed adaptive notch filter and the corresponding parameter estimation algorithm. This algorithm is a version of the recursive prediction error (RPE) system identification method of Ljung [5], [6]. The analysis of [6] has shown that this family of algorithms achieves the Cramer-Rao bound (CRB) asymptotically under weak assumptions in the minimal parameterization and model complete case (i.e., when the model set of the algorithm is minimal and

Manuscript received June 25, 1984; revised January 15, 1985.
The author is with Systems Control Technology, Inc., Palo Alto, CA 94304.

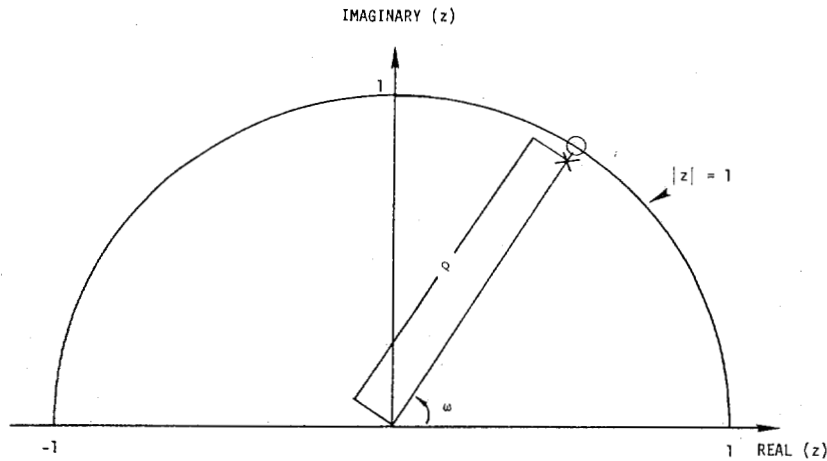


Fig. 1. Pole-zero configuration for a notch filter of a single sine wave in an additive broad-band process.

consistent with the data). For this reason, it is referred to as the recursive maximum-likelihood (RML) algorithm.

In Section III we analyze the performance of the proposed algorithm and give specific results for its accuracy in estimating the central frequency of a narrow-band autoregressive moving average (ARMA) process. We also present simulation results from Monte Carlo experiments for pure sine waves in white noise. The results show that the estimates of the sine wave frequencies obtained by the algorithm are of the same order of magnitude as the CRB for sufficiently large data sequences. Statistical results of this form (i.e., demonstrating the ability to approach the

where z is a complex variable and n corresponds to the number of notches. Thus, we shall concentrate on IIR notch filters with symmetric numerators as in (1).

The second requirement of the notch filter is that its poles be on the same radial lines as the zeros, but slightly displaced towards the origin. This can be achieved using filter denominators of the general form $A(\rho z^{-1})$, where ρ is a positive real number close to but smaller than 1. To see this, note that if $z^{2n}A(z^{-1})$ has a zero in z_i , then $z^{2n}A(\rho z^{-1})$ will have zero in ρz_i . By virtue of the foregoing arguments, we shall consider IIR notch filters in the general set

$$H(z^{-1}) = \frac{A(z^{-1})}{A(\rho z^{-1})} = \frac{1 + a_1 z^{-1} + \cdots + a_n z^{-n} + \cdots + a_1 z^{-2n+1} + z^{-2n}}{1 + \rho a_1 z^{-1} + \cdots + \rho^n a_n z^{-n} + \cdots + \rho^{2n-1} a_1 z^{-2n+1} + \rho^{2n} z^{-2n}} \quad (2)$$

CRB) were not previously presented in papers on adaptive filtering of sine waves in additive white noise. Section IV concludes the paper.

II. THE RECURSIVE ALGORITHM

A. The Proposed Model and Its Advantages

In this section we shall discuss the special structure of the proposed adaptive notch filter and derive its recursive parameter estimation algorithm.

As mentioned in the introduction, a desirable property of the notch filter is that the zeros of its transfer function will lie on the unit circle. A necessary condition for a polynomial to satisfy this property is that its coefficients will have a mirror symmetric form. A monic mirror symmetric polynomial can be written explicitly.¹

$$A(z^{-1}) = 1 + a_1 z^{-1} + \cdots + a_n z^{-n} + \cdots + a_1 z^{-2n+1} + z^{-2n} \quad (1)$$

¹Note that, in addition to the zeros on the unit circle, $A(z^{-1})$ also has $2n$ poles in the origin that are cancelled out by the similar poles of the denominator in the transfer function appearing in (2).

Let $y(t)$ be the observed time series at time t . The error output of the algorithm will be

$$\epsilon(t) = \frac{A(q^{-1})}{A(\rho q^{-1})} y(t) \quad (3a)$$

where q^{-1} is the unit delay operator, i.e., $q^{-1}y(t) = y(t-1)$, etc. The adaptive estimation algorithm will adjust the coefficients a_i so as to minimize the cost function

$$V_N = \sum_{t=1}^N \epsilon^2(t). \quad (3b)$$

Note that [1] and [3] also constrained the poles to lie on the same radial lines of the zeros. However, in contrast to these references, the zeros of $H(z^{-1})$ in (2) will be constrained to lie on the unit circle. In addition, notice that this transfer function, which is useful in eliminating n sine waves, is parameterized by only n coefficients. This is half the number of parameters in other adaptive notch algorithms of direct IIR type. The effect of the mirror symmetric constraint, in addition to the pole-zero coupling, is to limit the searched transfer functions to a minimal subset of all possible filters.

The model set above in (2) can also be explained as follows. Assuming that $y(t)$ consists of a sum of n sine waves plus white noise, say $v(t)$, it is well known (see, for example, [15], p. 1402) that $y(t)$ satisfies

$$A(q^{-1})y(t) = A(q^{-1})v(t) \quad (4)$$

where $A(q^{-1})$ is a $2n$ -order polynomial whose zeros lie on the unit circle at the sine wave frequencies [thus, $A(q^{-1})$ is symmetric and $y(t) \neq v(t)$]. Comparing (2) and (3a) with (4), one finds that under the assumption that $\rho \rightarrow 1$, the transfer function (2) corresponds to a proper and minimal whitening filter of $y(t)$, and $\epsilon(t)$ approximates the noise $v(t)$. The variable ρ is introduced for stability.

Observe that the mirror symmetric form of $A(z^{-1})$ is not sufficient to guarantee that the zeros of $H(z^{-1})$ will be on the unit circle. In fact, it only guarantees that if z_i is a zero of $H(z^{-1})$, then $1/z_i$ will also be a zero. However, to minimize the squared error function (3b), the convergent filter must place the zeros on the unit circle.

In the next section, we shall replace the constant pole moduli by a time-varying function. The use of a minimal and correct model (see above), as well as time-varying pole moduli, leads to several advantages of the proposed algorithm in comparison to previous algorithms, part of which were mentioned in the introduction. They are now explained in more detail as follows.

1) *Accuracy*: The proposed algorithm is expected to approach the CRB for sufficiently large data lengths by the convergence analysis of the RPE algorithms. Since previous adaptive notch filters were not minimally parameterized, they could not achieve the CRB (see also [6]). Another explanation for the better accuracy of the algorithm is by the parsimony principle [16]. By this principle, among RPE algorithms using hierarchical models (i.e., one can be obtained from the other by applying some constraints), the most accurate asymptotically is the one with the simplest model.

2) *Numerical Robustness*: Our application of user-controlled pole moduli rather than soft-constrained poles, such as in [4], can prevent numerical problems due to near pole-zero cancellations on the unit circle.

3) *Stability*: The special model used by our algorithm exhibits high stability, which is different from previous ANF algorithms and RPE in general. This saves computations needed to monitor the stability of RPE algorithms and enables the use of poles highly close to the unit circle, which result in significantly sharper notches than in previous algorithms (also, see below).

4) *Fast Convergence*: The relatively faster convergence of the algorithm below is due to three primary reasons. First, by the reduction of the dimension of the searched filter space to a minimal one. Second, by the coupling of the numerator coefficients to the denominator which increases the degree of linearity of $\epsilon(t)$ in the estimated parameter vector. (This remark is especially in comparison to general ARMA system identification al-

gorithms, for which it is known that their zero estimates converge more slowly than their pole estimates, in particular when the zeros are close to the unit circle.) The third reason for the fast convergence is related to the application of a special time-varying pole moduli (see, also, the next sections). The fourth reason for fast convergence is our use of Gauss-Newton type algorithms rather than stochastic gradient.

5) *Notch Filter Performance*: The advantages of using the algorithm as a notch filter are tied to the other advantages. The high accuracy of the sine wave frequency estimates is important to guarantee that the narrow notches will cancel the input sine waves. The bandwidth of the complex notches created by each pole-zero pair of our algorithm is given by

$$B = \pi(1 - \rho). \quad (5)$$

As will be shown in the next section, our algorithm will enable the use of any desirable value of ρ , and therefore, getting any arbitrary notch bandwidth. Since the zeros of the algorithms in [1] through [3] usually converged to a modulus less than one (see, for example, [1]), say r , their notch bandwidths were given by $B = \pi(1 - \rho r)$, which is larger than in our algorithm. The narrower bandwidth of our algorithm is important in order to obtain a smaller distortion in the wide-band component by the filter. In comparison to [4], that algorithm could not completely control the location of the poles, had nonuniform and wider notches than in ours, and required an extra notch at the Nyquist frequency.

6) *Computational Efficiency*: Recursive system identification algorithms of Gauss-Newton type require an amount of multiplications proportional to the squared number of estimated parameters at each time sample. Thus, our algorithm will need about a quarter of the computations of previous ANF algorithms.

7) *Linear Phase*: By the symmetry of the numerator in (2) and almost symmetry of the denominator, the convergent filter can be characterized as being close to linear phase (see, also, [7]).

B. Derivation of the Algorithm

We now turn to the derivation of the Gauss-Newton type RRE algorithm for the adaptive notch filtering. For this purpose, we shall first derive the associated regression expression of the predicted error $\epsilon(t)$ and its derivative w.r.t. the parameter vector defined by

$$\theta = [a_1 \cdots a_n]^T \quad (6)$$

where the superscript T denotes transpose. From (3a) it is possible to write the following difference equation:

$$\epsilon(t) = y(t) + y(t - 2n) - \rho^{2n}\epsilon(t - 2n) - \varphi^T(t)\theta \quad (7a)$$

where

$$\varphi(t) = [\varphi_1(t) \cdots \varphi_n(t)]^T \quad (7b)$$

and

$$\varphi_i(t) = \begin{cases} -y(t-i) - y(t-2n+i) + \rho^i \epsilon(t-i) \\ \quad + \rho^{2n-i} \epsilon(t-2n+i) & 1 \leq i \leq n-1 \\ -y(t-n) + \rho^n \epsilon(t-n) & i = n. \end{cases} \quad (7c)$$

Note that $\epsilon(t)$, $\varphi_i(t)$, and $\psi_i(t)$ below are in fact functions of θ , in addition to t .

The gradient of $\epsilon(t)$ w.r.t. the model θ is found as follows. Let

$$\psi(t) = [\psi_1(t) \cdots \psi_n(t)]^T \quad (8a)$$

where

$$\psi_i(t) = -\frac{\partial \epsilon(t)}{\partial a_i}. \quad (8b)$$

From (3a)

$$A(\rho q^{-1}) \epsilon(t) = A(q^{-1}) y(t). \quad (9)$$

Differentiating both sides of (9) w.r.t. a_i , one obtains, for $1 \leq i \leq n-1$,

$$\begin{aligned} A(\rho q^{-1}) \frac{\partial \epsilon(t)}{\partial a_i} + \rho^i \epsilon(t-i) + \rho^{2n-i} \epsilon(t-2n+i) \\ = y(t-i) + y(t-2n+i) \end{aligned} \quad (10a)$$

and for $i = n$,

$$A(\rho q^{-1}) \frac{\partial \epsilon(t)}{\partial a_n} + \rho^n \epsilon(t-n) = y(t-n). \quad (10b)$$

From these expressions and with (7b) and (7c), we find the relationship

$$\psi(t) = \frac{\varphi(t)}{A(\rho q^{-1})}. \quad (11)$$

This result implies that, except for the filtering through $1/A(\rho q^{-1})$, the gradient vector $\psi(t)$ has the same form as the regression $\varphi(t)$. Such filtering is common in RPE identification schemes (see, for example, [6], [8], and [9]). They are known to improve the convergence behavior of the algorithms. When applying the algorithm, however, the filter $1/A(\rho q^{-1})$ is unknown, and one has to replace it by its latest available estimate $1/\hat{A}(\rho q^{-1}, t)$. We will find it is convenient to define the filtered variables

$$y_F(t) = \frac{y(t)}{\hat{A}(\rho q^{-1}, t)} \quad (12a)$$

$$\epsilon_F(t) = \frac{\epsilon(t)}{\hat{A}(\rho q^{-1}, t)}. \quad (12b)$$

The RML algorithm for adaptive notch filtering, based on the above special modeling, is summarized below. To preserve the continuity, we chose to describe the special features of the algorithm after the equations rather than in between them.

C. The RML Adaptive Notch Filter

Design Variables: $n, \lambda(1), \lambda_0, \rho(1), \rho_0, \rho(\infty), \sigma$.

Initialization: $\hat{\theta}(0) = 0, P(0) = \sigma I_n, \psi(0) = \varphi(0) = 0, y(-i) = 0, i = 1, \dots, 2n$.

Nominal Values:

n = number of notches = number of parameters

$\sigma \cong 100/\bar{E}y^2(t)$

$\lambda(1) = 0.95 \quad \lambda_0 = 0.99$

$\rho(1) = 0.8, \rho_0 = 0.99, \rho(\infty) = 0.995$.

Main Loop:

$$\begin{aligned} \epsilon(t) = y(t) + y(t-2n) - \rho^{2n}(t) \bar{\epsilon}(t-2n) \\ - \varphi^T(t) \hat{\theta}(t-1) \end{aligned} \quad (13a)$$

$$\begin{aligned} P(t) = [P(t-1) - P(t-1) \psi(t) \psi^T(t) \\ \cdot P(t-1)/(\lambda(t) + \psi^T(t) P(t-1) \psi(t))]/\lambda(t) \end{aligned} \quad (13b)$$

$$\hat{\theta}(t) = \hat{\theta}(t-1) + P(t) \psi(t) \epsilon(t) \quad (13c)$$

$$\begin{aligned} \bar{\epsilon}(t) = y(t) + y(t-2n) - \rho^{2n}(t) \bar{\epsilon}(t-2n) \\ - \varphi^T(t) \hat{\theta}(t) \end{aligned} \quad (13d)$$

$$\begin{aligned} \bar{\epsilon}_F(t) = \bar{\epsilon}(t) - \rho^{2n}(t) \bar{\epsilon}_F(t-2n) - \sum_{i=1}^{n-1} [\rho^i(t) \bar{\epsilon}_F(t-i) \\ + \rho^{2n-i}(t) \bar{\epsilon}_F(t-2n+i)] \hat{a}_i(t) \\ - \rho^n(t) \bar{\epsilon}_F(t-n) \hat{a}_n(t) \end{aligned} \quad (13e)$$

$$\begin{aligned} y_F(t) = y(t) - \rho^{2n}(t) y_F(t-2n) - \sum_{i=1}^{n-1} [\rho^i(t) y_F(t-i) \\ + \rho^{2n-i}(t) y_F(t-2n+i)] \\ \cdot \hat{a}_i(t) - \rho^n(t) y_F(t-n) \hat{a}_n(t) \end{aligned} \quad (13f)$$

$$\varphi_i(t) = \begin{cases} -y(t-i) - y(t-2n+i) \\ \quad + \rho^i(t) \bar{\epsilon}(t-i) + \rho^{2n-i}(t) \bar{\epsilon}(t-2n+i) & 1 \leq i \leq n-1 \\ -y(t-n) + \rho^n(t) \bar{\epsilon}(t-n) & i = n \end{cases} \quad (13g)$$

$$\psi_i(t) = \begin{cases} -y_F(t-i) - y_F(t-2n+i) \\ \quad + \rho^i(t) \bar{\epsilon}_F(t-i) + \rho^{2n-i}(t) \\ \quad \cdot \bar{\epsilon}_F(t-2n+i) & 1 \leq i \leq n-1 \\ -y_F(t-n) + \rho^n(t) \bar{\epsilon}_F(t-n) & i = n \end{cases} \quad (13h)$$

$$\varphi(t+1) = [\varphi_1(t+1) \cdots \varphi_n(t+1)]^T \quad (13i)$$

$$\psi(t+1) = [\psi_1(t+1) \cdots \psi_n(t+1)]^T \quad (13j)$$

$$\lambda(t+1) = \lambda_0 \lambda(t) + (1 - \lambda_0) \quad (13k)$$

$$\rho(t+1) = \rho_0 \rho(t) + (1 - \rho_0) \rho(\infty). \quad (13l)$$

D. Special Features and General Remarks

Time-Varying ρ : Notice that in the foregoing recursions, the previous fixed modulus ρ of the poles was replaced by a time-varying function $\rho(t)$. This is primarily motivated according to the following two reasons. Recall that ρ determines the bandwidth of each of the notches. In practical cases, if no *a priori* information is available on the input sine waves, and if the notches are too narrow, the notches may not fall over the sine wave frequencies and the algorithm will not "sense" the presence of the sine waves. This may prevent the algorithm from converging to the desired transfer function. Therefore, at the start of the data processing, it is advisable to apply the algorithm with wider notches (i.e., smaller values of ρ), thus increasing filter sensitivity to the presence of input sine waves. After convergence, it is recommended that a larger modulus of poles (closer to 1) be used, which improves the asymptotic performance of the algorithm.

The second reason for using a time-varying modulus of the poles is that at the beginning of the algorithm operation, the estimate of $A(q^{-1})$ is usually not good. This may cause a poor filtering in (13e) and (13f). The effect of this problem may be diminished by decreasing the decay time of the filter, thus "weakening" its effect for moderate values of t . This can be achieved by moving the poles of $1/\hat{A}(\rho q^{-1}, t)$ closer to the origin at the start of the data processing.

For the above two reasons, it is useful to apply the algorithm with small values of ρ at the beginning of the algorithm's operation, and increase its value later. A simple way to do this is to let ρ grow exponentially from a desired value $\rho(1)$ to $\rho(\infty)$ according to

$$\rho(t+1) = \rho_0 \rho(t) + (1 - \rho_0) \rho(\infty). \quad (14a)$$

Notice that the parameter ρ_0 determines the rate of changes in $\rho(t)$. For values of ρ_0 close to one, which is usually the case, ρ_0 corresponds to an exponential decay time constant or "memory length" given by

$$T_\rho = \frac{1}{1 - \rho_0}. \quad (14b)$$

In our simulation studies, we applied the algorithm with the numerical values

$$\rho(1) = 0.8, \quad \rho_0 = 0.99, \quad \rho(\infty) = 0.995.$$

The value $\rho(1) = 0.8$ was chosen as the smallest pole modulus which has a significant effect on the bandwidth of the notch filter. The value $\rho(\infty) = 0.995$ was found to yield most accurate results in estimating the sine wave frequencies in the Monte Carlo experiments (see the next section). However, it is interesting to remark that the algorithm could also work well with values of $\rho(\infty)$ extremely close to 1, namely $\rho(\infty) = 1 - \epsilon$, where ϵ is an arbitrary small number whose value is limited only by the accuracy of the computer (sometimes called "machine epsilon"). For instance, in our double-precision computa-

tions, ϵ had the value $\epsilon = 1.39 \times 10^{17}$. This means that this algorithm can achieve practically any prescribed notch bandwidth, as narrow as desired!

The algorithm's ability to work with the modulus of the poles extremely close to the unit circle is remarkable from a stability standpoint, as is discussed next. Previous adaptive notch filter algorithms were incapable of providing comparable highly narrow bandwidth notches.

Notice that $\rho(t)$ plays a similar role to α_i in [4] in controlling the decay time of the prefiltering process. However, $\rho(t)$ also has the important function of controlling the notch widths, which could not be done with α_i in [4].

Stability Monitoring: The algorithm derivation and convergence analysis by Lung in [5], [6], and [10] has shown that stability monitoring must be incorporated in the RPE class of algorithms. This means that at each time t , one must check to see if $\hat{\theta}(t)$ corresponds to a stable filter, and if not, to project it back into the stable model set in order to prevent divergence. In our case, where the poles are extremely close to the unit circle, one may suspect that the stability projection has to be applied very often. In fact, in [2], p. 804 and 806 noted such effects of instability in the adaptive notch filter, and in [4], p. 291 reported "heavy use of angle-invariant (stability) projections" in attempting to place the poles close to the unit circle. (It should be noted that the use of a simpler algorithm of LMS type could be part of the reason for instability in [2].) Interestingly enough, even though our poles are much closer to the unit circle, we found in all our extensive simulations that the present algorithm is remarkably stable, and the stability monitoring was not applied at all!

This result was tested and verified for multiple sine waves and values for $\rho(\infty)$ equal to $1 - \epsilon$, where ϵ is the "machine epsilon" (see above). Clearly, it is of practical value in situations where the computational burden of the stability monitoring is not affordable. The excellent stability of the new algorithm is undoubtedly due to the special constraints imposed on the filter. Thus, note that our use of a symmetric polynomial $A(q^{-1})$ as well as pole-zero contraction factor $\rho < 1$ yields a dense set of θ values which corresponds to stable $H(q^{-1})$. In addition, our use of special time-varying pole modulus in (131) increases the stability robustness, especially at the transient phase.

Updating $P(t)$ and $\lambda(t)$: The recursions (13b) and (13k) for $P(t)$ and $\lambda(t)$, respectively, are adapted from [6] for stationary signals. Note that with these recursions $P(t) = \bar{R}^{-1}(t)$, where $\bar{R}(t) = \lambda(t) \bar{R}(t-1) + \psi(t) \psi^T(t)$. The matrix $\bar{R}(t)$ approximates the second derivative matrix of the cost function w.r.t. θ evaluated at $\hat{\theta}(t-1)$. For this reason, this algorithm can be viewed as a recursive Gauss-Newton type.

When tracking nonstationary parameters, it is common to choose $\lambda(t) \equiv \lambda_0 < 1$. The effect of using a fixed value λ_0 is the modification of the cost function (3b) to an exponentially weighted cost function.

A Posteriori Prediction Error: The recursion (13d) means that at each time t the prediction error

$$\epsilon(t) = y(t) - \hat{y}(t|\hat{\theta}(t-1)) \quad (15a)$$

is replaced by the residual or a *posteriori* prediction error

$$\bar{\epsilon}(t) = y(t) - \hat{y}(t|\hat{\theta}(t)). \quad (15b)$$

This procedure was found to improve the convergence rate of system identification schemes (see, for example, [6]), since usually $\bar{\epsilon}(t)$ is expected to yield better estimates for the error than $\epsilon(t)$.

Convergence Analysis Results: The convergence analysis in [5], [6], and [10] for recursive identification algorithms, established specific results for the asymptotic properties of the RPE type algorithms. The assumptions under which this analysis holds are generally weak, and can be briefly summarized as:

1) stable generation of the data in the algorithm; 2) smooth parameterization; 3) bounded matrix $tP(t)$, and 4) gain $\gamma(t) = \gamma(t-1)/(\gamma(t-1) + \lambda(t))$ behaving asymptotically as $1/t$. Under these assumptions, it was shown that the asymptotic trajectories of the RPE parameter estimates coincide with those of an ordinary differential equation (ODE) and the RPE converges to a local minimum of the cost function (3b) or to the boundary of the model set. It was also shown that in the minimal parameterization, model complete and white Gaussian noise case, if the data are bounded and if the state matrices in the algorithm are twice differentiable, then the parameter estimates of the RPE are asymptotically unbiased, normally distributed, and achieve the Cramer-Rao bound (i.e., statistical efficiency).

To apply the above results to our algorithm, one has to note first that this algorithm is a special RPE algorithm. Second, it should be proven that all the necessary conditions in the convergence analysis summarized above hold here in the model complete case. We will skip the details and only note that they can be easily proved. Thus we conclude that the convergence analysis and the results by the ODE approach of the RPE algorithms are applicable here in the model complete case. In addition, we note that by using the ODE approach, it can be shown that, for sine waves in additive white noise, the only possible convergent point of the algorithm is the true parameter vector (i.e., consistency).

III. PERFORMANCE EVALUATION

This section is devoted to performance evaluation of the new adaptive notch filter algorithm presented in Section II. The results will also show how this algorithm can be modified for several related applications. We shall first analyze the expected accuracy of the algorithm in estimating the central frequency of a narrow-band ARMA process using the Cramer-Rao bound. Then we will present simulation results for multiple sine waves in white noise, including Monte Carlo experiments and comparison of the sine wave frequency estimates to the Cramer-Rao bound. At the last part, we shall evaluate the performance in applications of adaptive line enhancement and discuss a procedure for order determination.

A. The CRB for Narrow-Band ARMA Estimation

One of the important results of the RPE class of algorithms is their asymptotic efficiency, i.e., they achieve the

Cramer-Rao bound in the minimal parameterization and model complete case for sufficiently large data sets. The family of ARMA signals corresponding to our adaptive notch filter algorithm can be obtained by modifying the inverse transfer function of (2) and (3a) to a stable one. The resulting process is characterized by

$$y(t) = \frac{A(\rho q^{-1})}{A(rq^{-1})} e(t) \quad (16)$$

where $e(t)$ is a sequence of independent zero-mean random Gaussian variables with equal variance. The variables ρ and r satisfy the relation $\rho < r < 1$ and the associated polynomial $A(q^{-1})$ is symmetric as in (1). Note that when ρ and r are close to 1, the process (16) models multiple narrow-band ARMA signals in an efficient way. The adaptive notch filter of Section II can be easily modified to this model by replacing $y(t-i)$ with $r^i y(t-i)$.

By the general results of statistics, the covariance of an unbiased estimator of the true parameter vector θ_0 is bounded by

$$\text{cov } \hat{\theta} \geq J^{-1} \quad (17a)$$

where J is the $n \times n$ Fisher information matrix given by

$$J = -E \left[\frac{\partial^2}{\partial \theta_i \partial \theta_j} \ln p(Y|\theta_0) \right] \quad (17b)$$

and $p(Y|\theta_0)$ is the likelihood function (conditional probability distribution) of the observed data Y .

To demonstrate the performance of our algorithm in estimating the parameters of the narrow-band ARMA processes (16), we shall consider a second-order case, i.e., a single narrow-band ARMA process modeled with the aforementioned constraints. Here

$$y(t) = \frac{1 + \rho a_1 q^{-1} + \rho^2 q^{-2}}{1 + r a_1 q^{-1} + r^2 q^{-2}} e(t). \quad (18)$$

Let

$$b = (\rho - r) [1 \ 0 \ -\rho r \ 0 \ 0]^T \quad (19a)$$

$$C(q^{-1}) = A(\rho q^{-1}) A(rq^{-1}). \quad (19b)$$

With these definitions, we find in the CRB analysis of the Appendix that the asymptotic error variance of a_1 is given by

$$\text{var}(\hat{a}_1) = \frac{1}{Nb^T [T_1 T_1^T - T_2 T_2^T]^{-1} b} \quad (20)$$

where T_1 and T_2 are 4×4 lower triangular Toeplitz matrices characterized by their first columns

$$T_{11} = [1 \ c_1 \ c_2 \ c_3 \ c_4]^T \quad (21a)$$

$$T_{21} = [0 \ c_4 \ c_3 \ c_2 \ c_1]^T \quad (21b)$$

and the variables c_i are the coefficients of $C(q^{-1})$ defined in (19b). From (20), a close formula for the asymptotic variance of \hat{f}_1 is derived in the Appendix. The result is

$$\text{var}(\hat{f}_1) = \frac{1}{4\pi^2 N (4 - a_1^2) b^T [T_1 T_1^T - T_2 T_2^T]^{-1} b}. \quad (22)$$

TABLE I
PER-SAMPLE CRAMER-RAO BOUND FOR A SINGLE NARROW-BAND ARMA
PROCESS. CENTRAL FREQUENCY $f_1 = 0.125$

PARAMETERS		CRAMER-RAO BOUND	
ρ	r	$\text{Var}(\hat{c}_1)$	$\text{Var}(\hat{f}_1)$
0.8944	0.98	0.0712	9.012×10^{-4}
0.9487	0.99	0.0366	4.629×10^{-4}
0.9747	0.995	0.0185	2.345×10^{-4}
0.9950	0.999	0.0037	0.477×10^{-4}

Numerical Illustration: Using the above expressions, we can calculate specific results for the asymptotic error variance of our algorithm in estimating the central frequency of a narrow-band ARMA process. The following example is adapted from [4]. It will demonstrate how previous signal models can be parameterized more efficiently according to our model.

The central frequency of the ARMA process in the example adapted from [4] is $f_1 = 0.125$. The values of r below are chosen to obtain zeros at modulus 0.98, 0.99, 0.995, and 0.999. (They are similar to [4]; note, however, that the zero moduli are directly parameterized here by r , which was not the case in [4].) The values of ρ are computed by a proper transformation of the signal model in [4]. Specifically, our ρ^2 equals c_2 of [4]. Observe that we have thus obtained an equivalent signal model to the one in [4] using only the single parameter a_1 , while [4] needed two parameters. The numerical results of the CRB obtained for the narrow-band ARMA signal are summarized in Table I.

Comparing the results of Table I to those of [4], we find that the asymptotic error variance of \hat{a}_1 in our algorithm is somewhat larger than the error of \hat{c}_1 in [4], and smaller than that of \hat{c}_2 in [4]. (Note that both \hat{c}_1 and \hat{c}_2 of [4] are related to \hat{a}_1 of our algorithm.) To obtain more conclusive results, it would be useful to compare the asymptotic error variances of the central frequency \hat{f}_1 . Unfortunately, this is not possible due to several errors in [4]. We should note, however, that under the assumption that ρ and r are fixed numbers determined by the ARMA process and not estimated parameters, the asymptotic error variance of our algorithm will be smaller due to the parsimony principle.

B. Simulation Results for Sine Waves in White Noise

In the examples below, the algorithm of Section II was implemented using a DEC VAX computer with double precision arithmetic. The nominal values of the design variables in (13) were used, in particular, $\rho(\infty) = 0.995$. Stability was assumed; therefore, stability monitoring was not applied in any of the simulation runs.

Monte Carlo Simulations: By the general convergence results of the RPE algorithms, the adaptive notch filter of Section II is expected to approach the Cramer-Rao bound in estimating the frequencies of sine waves in additive white Gaussian noise for sufficiently large data sets. The

following examples demonstrate this result by Monte Carlo experiments from which the sample statistics of the frequency estimates are computed. Each of the statistical results was computed from 40 independent experiments. The parameter estimates $\hat{\theta}(t)$ were set to zero at the beginning of each experiment. The algorithm was tested under different data lengths and signal-to-noise ratios.

In the first example, the input data were

$$y(t) = C_1 \sin 2\pi 0.1t + C_2 \sin 2\pi 0.2t + v(t) \quad (23)$$

where $v(t)$ is a zero-mean unit-variance white Gaussian noise. The prescribed SNR conditions were obtained by adjusting the amplitudes C_i according to $\text{SNR (dB)} = 10 \log (C_i^2/2)$. Table II summarizes the resulting sample statistics of the sine wave frequency estimates. These estimates were obtained by factoring the polynomial $\hat{A}(q^{-1}, t)$ at the end of each experiment. In all experiments, the zeros of this polynomial were precisely on the unit circle.

In the simulations, we encountered some realizations which gave outlier performance, i.e., the estimates were not close to the typical behavior. The number of outliers out of 40 experiments in each frequency estimate is indicated by superscript in the table. The corresponding experiments (i.e., the whole results for both frequency estimates) were eliminated from the statistical computations. This is justified by the fact that the Cramer-Rao bound is derived under small error assumption. The decision on outlier was based on an arbitrary threshold of 0.01 in the error of the frequency estimates, except for the case of two sine waves with $N = 2000$ and 4 dB, where 0.007 was used. Note that the number of cases in which this happened is relatively small and occurred only for moderate SNR's. Also, some of the outliers corresponded to results which may be acceptable in some practical applications, but they were excluded from the statistics to obtain a fair comparison with the CRB.

A close examination of the outlier behavior showed that the outliers were not due to double-root polynomial estimates. A trial to reduce the number of outliers by applying (unnecessary) stability monitoring was resulted by an opposite effect. Results for lower SNR's than in the tables showed that the number of outliers increased significantly, but the results which were not considered outliers remained highly accurate. We may also note that the zeros converged to the unit circle even in the outlier cases.

Table II also presents the minimal achievable standard deviation of the sine wave frequency estimates by the Cramer-Rao bound. This bound was computed using (square root of) the formula given by Rife and Boorstyn [12]

$$\text{var}(\hat{f}_i) = \frac{3}{\pi^2 N^3 \text{SNR}} \quad (24)$$

Equation (24) is applicable for well separated sine waves, where $|f_i - f_j| \gg 1/N$. Note that in the ARMA estimation, the variance (22) of the central frequency was proportional to $1/N$, while for pure sine waves, (24) is proportional to $1/N^3$. The results of the table show that the

TABLE II
STATISTICAL RESULTS FOR TWO SINE WAVES IN ADDITIVE WHITE NOISE

N	SNR (dB)	BIAS \hat{f}_1	ST. DEV. \hat{f}_1	BIAS \hat{f}_2	ST. DEV. \hat{f}_2	CRB
100		$\times 10^{-4}$	$\times 10^{-4}$	$\times 10^{-4}$	$\times 10^{-4}$	$\times 10^{-4}$
	0	-8.91 (3)	20.8 (3)	-3.33 (2)	23.8 (2)	5.51
	4	-6.06	17.2	-4.86	12.6	3.48
	8	-1.42	6.03	-0.04	8.30	2.19
	12	0.37	3.88	0.49	3.16	1.38
	16	1.31	1.90	-0.94	2.49	0.87
	20	1.44	1.56	-0.91	1.47	0.55
500		$\times 10^{-6}$	$\times 10^{-5}$	$\times 10^{-6}$	$\times 10^{-5}$	$\times 10^{-5}$
	0	-184.9 (2)	91.4 (2)	-179.8	140.6	4.93
	4	8.49	11.5	-6.99	13.5	3.11
	8	-0.90	8.09	-1.86	6.20	1.96
	12	-5.26	3.84	1.49	4.11	1.24
	16	1.59	3.05	-1.90	2.62	0.78
	20	1.91	1.94	3.29	1.93	0.49
2000		$\times 10^{-7}$	$\times 10^{-6}$	$\times 10^{-7}$	$\times 10^{-6}$	$\times 10^{-6}$
	0	5.46	11.9	64.8 (2)	22.7 (2)	6.16
	4	19.1 (1)	7.25 (1)	7.42 (1)	7.79 (1)	3.89
	8	-0.72	4.71	3.03	4.89	2.45
	12	-4.92	3.37	2.43	2.74	1.55
	16	5.10	2.34	-0.45	2.11	0.98
	20	-0.93	1.25	-0.95	1.09	0.62

sample standard deviations of the algorithm are of the same order of magnitude as the CRB for sufficiently large sample size N .

In the next example, the data were

$$y(t) = \sum_{i=1}^4 C_i \sin 2\pi 0.1it + v(t). \quad (25)$$

The statistical results are summarized in Table III. We observe that the accuracy of the frequency estimates is not affected by the number of sine waves.

Convergence Illustrations: The following examples demonstrate the convergence of the algorithm by illustrations of two typical simulation runs. Fig. 2(a) depicts the parameter estimates \hat{a}_1 and \hat{a}_2 versus time for the input (23) and SNR = 0 dB. Fig. 2(b) shows the resultant pole-zero diagram of the convergent filter (at $t = 1000$). Similar results could be obtained for a smaller t . As desired, the convergent zeros are precisely on the unit circle at the sine wave frequencies and the convergent poles are on the same radial lines as the zeros with a modulus equal to

0.995. Fig. 2(c) illustrates the magnitude of the corresponding transfer function. Observe that the notches are highly narrow and uniform as should be expected from our special constraints imposed on the filter. Note that even narrower notches could be obtained by using larger values of $\rho(\infty)$.

Fig. 3 illustrates similar results for the signal

$$y(t) = \sum_{i=1}^5 C_i \sin 2\pi 0.08it + v(t) \quad (26)$$

and SNR = 0 dB. It can be seen that in this higher order, the convergence is slower than in the previous case. When applying the algorithm in higher signal-to-noise ratios, the convergence was faster and smoother. Such results are common in system identification schemes (see, for example, [8]).

Comparison to Previous Results: Since the advantages of the proposed algorithm in comparison to previous algorithms were mentioned above, they will not be repeated here. The simulation results, however, emphasize some of

TABLE III
STATISTICAL RESULTS FOR FOUR SINE WAVES IN ADDITIVE WHITE NOISE

N	SNR (dB)	BIAS \hat{f}_1	ST. DEV. \hat{f}_1	BIAS \hat{f}_2	ST. DEV. \hat{f}_2	BIAS \hat{f}_3	ST. DEV. \hat{f}_3	BIAS \hat{f}_4	ST. DEV. \hat{f}_4	CRB
100		$\times 10^{-5}$	$\times 10^{-4}$	$\times 10^{-5}$	$\times 10^{-4}$	$\times 10^{-5}$	$\times 10^{-4}$	$\times 10^{-5}$	$\times 10^{-4}$	$\times 10^{-4}$
	4	-116.5 (1)	26.1 (1)	-4.05 (2)	9.54 (2)	17.1 (1)	8.11 (1)	10.0 (1)	9.20 (1)	3.48
	8	-13.6	5.43	-15.8	3.89	-1.12	4.84	3.78	4.99	2.19
	12	-9.14	2.69	0.66	3.12	8.28	2.57	8.21	2.79	1.38
	16	0.62	1.81	2.59	1.56	0.88	2.09	-0.14	1.69	0.87
	20	0.26	0.96	-3.00	1.25	0.53	1.14	-1.36	1.10	0.55
500		$\times 10^{-6}$	$\times 10^{-5}$	$\times 10^{-6}$	$\times 10^{-5}$	$\times 10^{-6}$	$\times 10^{-5}$	$\times 10^{-6}$	$\times 10^{-5}$	$\times 10^{-5}$
	4	5.68 (2)	10.2 (2)	15.6 (2)	11.5 (2)	2.35 (3)	11.0 (3)	-19.1 (1)	8.51 (1)	3.11
	8	-1.57	6.66	6.79	7.53	2.87	6.82	-0.97	5.92	1.96
	12	2.68	4.19	-4.06	5.04	6.31	3.73	-6.02	3.86	1.24
	16	-0.71	2.93	-6.82	2.23	-3.14	2.95	5.28	2.77	0.78
	20	-2.32	1.36	0.58	1.53	-1.55	1.28	1.49	1.35	0.49
2000		$\times 10^{-7}$	$\times 10^{-6}$	$\times 10^{-7}$	$\times 10^{-6}$	$\times 10^{-7}$	$\times 10^{-6}$	$\times 10^{-7}$	$\times 10^{-6}$	$\times 10^{-6}$
	4	-5.55	6.75	20.6 (3)	5.58 (3)	-2.04 (2)	8.01 (2)	-13.3	7.15	3.89
	8	10.3	4.26	3.65	4.80	-21.7	4.59	-9.41	4.61	2.45
	12	0.46	2.74	1.33	2.94	-7.74	2.71	-1.08	2.60	1.55
	16	-3.33	2.01	1.42	1.67	0.77	1.60	-3.88	1.78	0.98
	20	3.06	1.39	-0.25	1.27	-4.61	1.21	-4.69	1.16	0.62

the advantages and we note the following. Recall that our algorithm did not need stability projections in any of the simulations. In contrast, the algorithm of [4] required "heavy use of angle invariant (stability) projections" (see [4], p. 291) when soft constraints were imposed in attempts to place the poles close to the unit circle. (Note that our poles are much closer to the unit circle.)

As noted in [4], p. 291, for two sine waves and SNR = 0 dB, the algorithm of [4] did not completely converge even after 2048 data points. This is very different from our simulation results, which show convergence after about 50 to 70 data points for similar conditions [see Fig. 2(a)]. A comparison of the convergent transfer functions [see Fig. 2(c) and 3(c)] shows that the notches in our algorithm are completely uniform and much narrower than in [4] (see Figs. 4 and 5 of [4]). Clearly, this result is due to the special constraints of our filter which dictate the shape of the notches and the resulting stability which enables the use of poles highly close to the unit circle. In addition, note that the algorithm in [4] necessarily had a superfluous notch at the Nyquist frequency, while the algorithm proposed here does not.

Results from Monte Carlo simulations were recently published in [3] (see Tables I-IV) using adaptive notch filter based on the earlier reports [1], [2]. A comparison of these results with ours shows that the convergent zeros in these references has modulus smaller than one, while ours has modulus precisely equal to one. Also, the pole-zero contraction factor yielding most accurate results in [3] is 0.9, while ours is 0.995. The use of this larger value of pole-zero contraction factor (and the other differences mentioned above) has resulted in more accurate results and has enabled our algorithm to approach the CRB, as was illustrated by the computer simulations. Results of this type were not presented before.

It is also of interest to note that the parameter estimates in the adaptive notch filter of this paper exhibit smoother and faster convergence than those arising in general ARMA system identification schemes (see, for example, [8]). One explanation for the difference in smoothness is due to the different powers in N between (22) and (24).

Sine Waves with Different SNR's: In addition to the above examples, we have also simulated cases of two sine waves with different SNR's. The two sine waves had char-

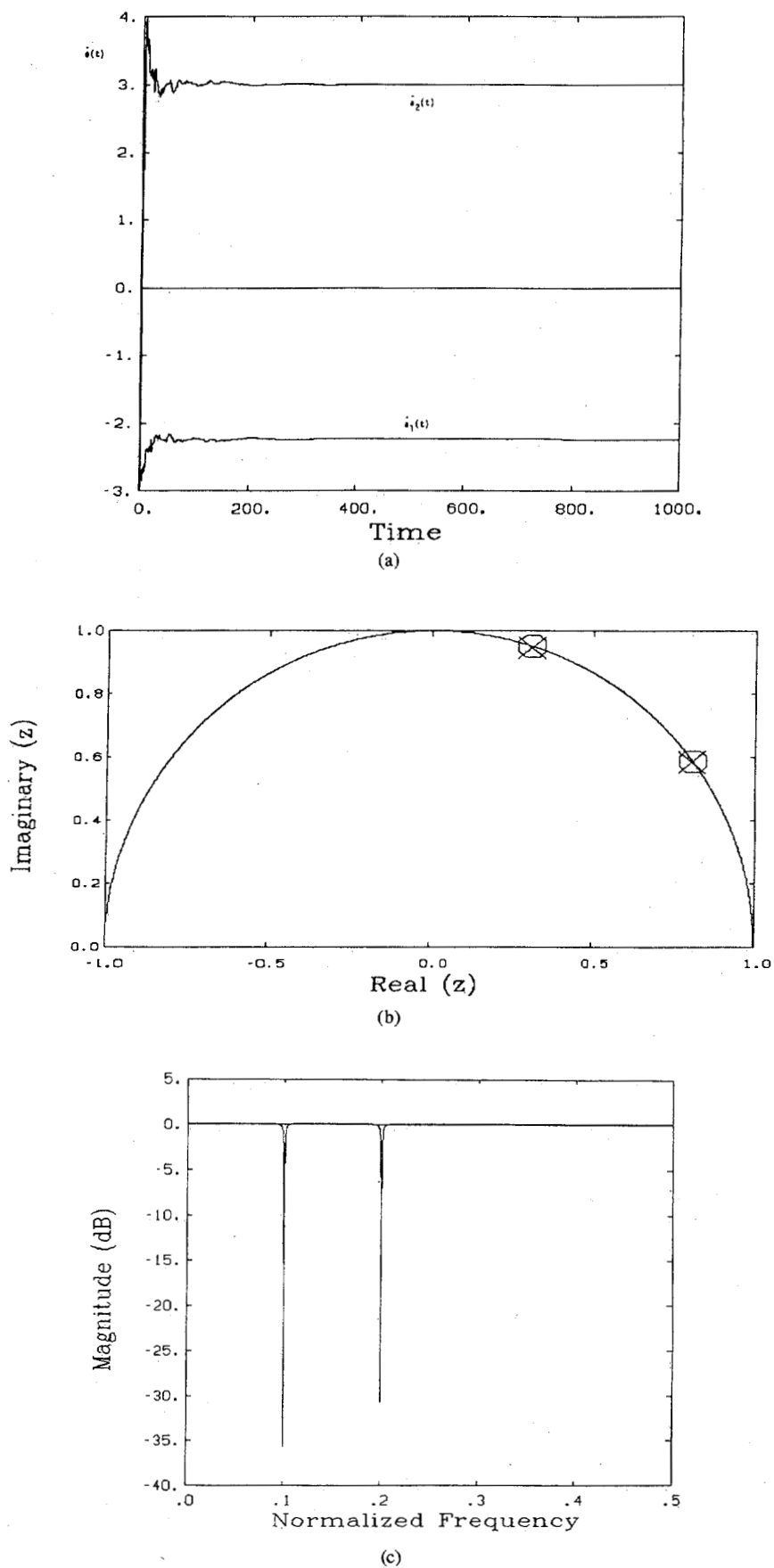


Fig. 2. Estimation results for two sine waves in additive white noise. SNR = 0 dB. (a) Parameter-estimate trajectories. (b) Pole-zero diagram of the convergent filter. (c) Transfer function of the convergent filter.

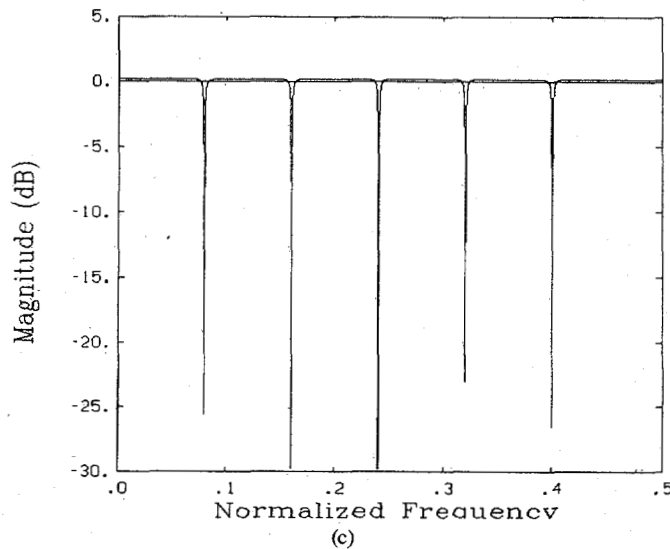
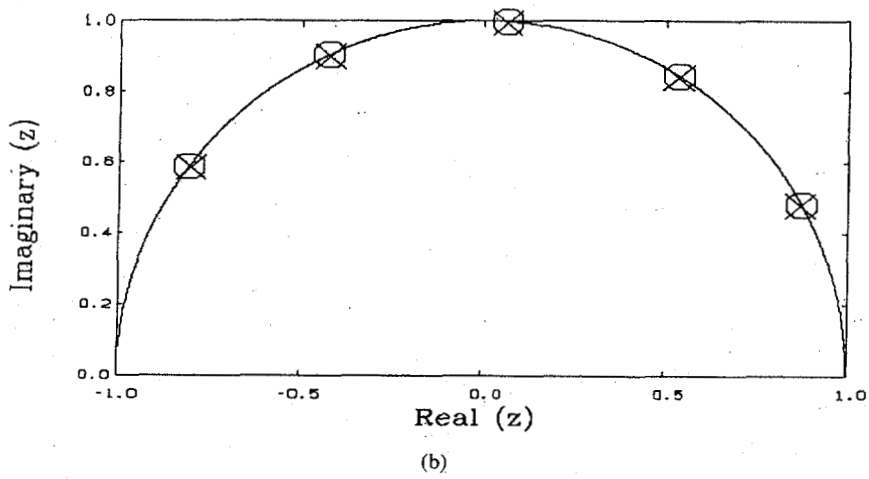
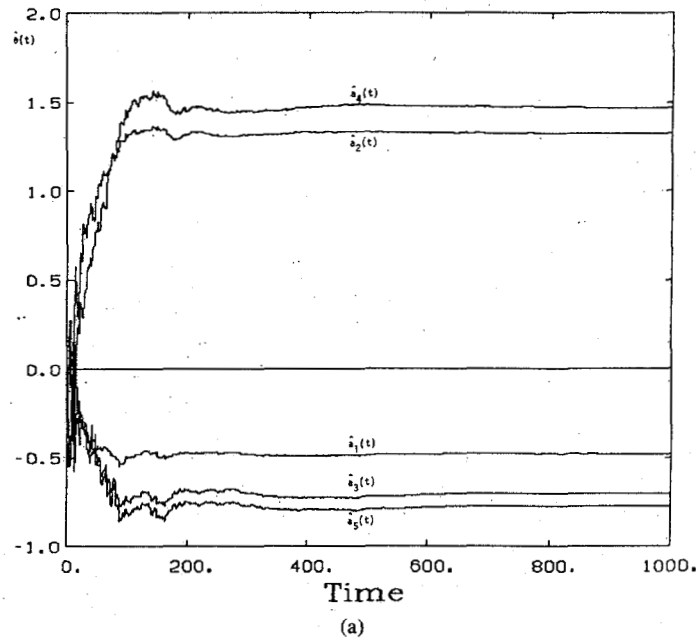


Fig. 3. Estimation results for five sine waves in additive white noise. SNR = 0 dB. (a) Parameter-estimate trajectories. (b) Pole-zero diagram of the convergent filter. (c) Transfer function of the convergent filter.

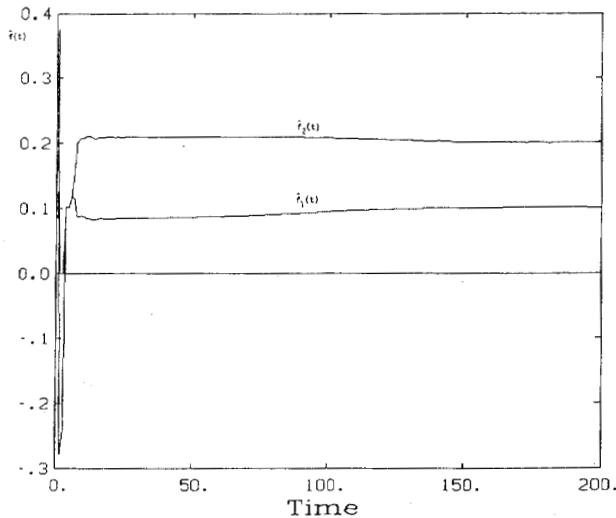


Fig. 4. Estimation results for two sine waves with different powers. $\text{SNR}_1 = 20$ dB, $\text{SNR}_2 = -4$ dB.

acteristics as in (23), and an SNR difference of 8 dB. Results from Monte Carlo simulations for this type of input have shown that the frequency estimates still approached the CRB despite the different SNR's for large data lengths of $N = 500$ and $N = 2000$. However, for $N = 100$, the results were deteriorated in comparison to the above examples of equal SNR sine waves.

One of the well-known drawbacks of the stochastic gradient adaptive line enhancer (ALE) in [13] is the different convergence rate for sine waves with different SNR's. This drawback usually does not characterize system identification algorithms such as ours, which employ Gauss-Newton type recursions. To demonstrate this feature, we simulated examples of two sine waves in additive white noise, characterized by $f_1 = 0.1$, $f_2 = 0.2$, $\text{SNR}_1 = 20$ dB, and $\text{SNR}_2 = -4$ dB. Fig. 4 depicts the frequency estimates for the two sine waves as a function of time. Note that the convergence rate is about the same for the two sine waves, despite the large difference in their SNR's.

Results for Overdetermined Order: Application of the algorithm with an overdetermined order showed that in such cases the algorithm requires stability monitoring. In addition, the results are, in general, less good than in the correct order case. This is in agreement with our accuracy claim above due to the minimal parameterization. We may note that all convergent zeros lay on the unit circle.

C. Line Enhancement and Order Determination

Before concluding the paper, we shall briefly comment on two more possible applications of the proposed adaptive notch filter algorithm, including adaptive spectral line enhancement and order estimation.

Line Enhancement: In some applications, the purpose is to extract the narrow-band components from observations corrupted by noise. It is easy to see that this can be done with the above adaptive notch filter simply by subtracting the error output $\epsilon(t)$ from the input $y(t)$. For input of multiple sine waves in white noise, the asymptotic sig-

nal-to-noise ratio gain obtained in this way can be shown to be proportional to the inverse notch filter bandwidth. Specifically, in our algorithm

$$\frac{\text{SNR}_0}{\text{SNR}_i} = \frac{1}{1 - \rho} \quad (27)$$

In comparison, the algorithms of [1] through [3] have gain [1]

$$\frac{\text{SNR}_0}{\text{SNR}_i} = \frac{1}{1 - \rho r} \quad (28)$$

where r denotes the modules of the zeros in [1] through [3]. Since these zeros were not constrained to lie on the unit circle, r is less than 1, and our gain (27) is larger than (28).

Another algorithm which is more widely known for spectral line enhancement is the stochastic gradient-type adaptive line enhancer (ALE) of [13]. This algorithm is based on AR modeling of the input, which results in a finite impulse response (FIR) adaptive filter. The optimal SNR gain of this filter for equal power sine waves was found to be [14]

$$\frac{\text{SNR}_0}{\text{SNR}_i} = L/(2N) \quad (29)$$

where L is the number of the FIR coefficients and N is the number of input sine waves. This result implies that the FIR-ALE filter of [13] usually needs a relatively large number of coefficients to achieve comparable performance to ours. For instance, to obtain SNR gain of $1/(1 - 0.995) = 200$, which is typical of our IIR adaptive notch filter algorithm, the FIR-ALE needs 400 coefficients per sine wave! Clearly, this difference is due to the different model in [13].

Order Determination: For the sake of completeness, we shall present one possible way of estimating the signal order, using the following steps. First, factorize the estimated filter $\hat{A}(q^{-1})$ to obtain the frequency estimates \hat{f}_i of the input narrow-band components. Next, filter the measurements $y(t)$ through the resonant bandpass filters

$$\hat{H}_i(q^{-1}) = 1 - \frac{\hat{A}_i(rq^{-1})}{\hat{A}_i(\rho q^{-1})} \quad 1 \leq i \leq n \quad (30)$$

where

$$\begin{aligned} \hat{A}_i(q^{-1}) &= 1 + \hat{a}_{i1}q^{-1} + q^{-2} \\ \hat{a}_{i1} &= -2 \cos 2\pi\hat{f}_i \end{aligned}$$

and $\rho < r \leq 1$. The output of each of these filters yields an estimate of the i th narrow-band component of the input, whose central frequency is \hat{f}_i . Thus, we write

$$\hat{x}_i(t) = \left[1 - \frac{\hat{A}_i(rq^{-1})}{\hat{A}_i(\rho q^{-1})} \right] y(t) \quad (31)$$

The order of the input signal can now be estimated by using a sufficiently large order and measuring the powers of $\hat{x}_i(t)$. If the power of some of these signals is less than

a predetermined threshold, the algorithm order has to be reduced by twice the number of these low-energy signals. This method can be applied either in an on-line or an off-line version.

IV. CONCLUSION

A new algorithm has been presented for adaptive notch filtering and parametric spectral estimation of narrow-band or sine wave signals in an additive broad-band process. It has been shown that the variances of the estimated frequencies of the narrow-band or sine wave signals are of the same order of magnitude as the Cramer-Rao bound for sufficiently large data sets. This is the first adaptive filter for which such results are presented. The convergent filter is characterized by highly narrow and uniform notches of desirable shape, yet this algorithm requires a relatively small amount of computation.

Simulation results have illustrated that the proposed algorithm is consistently stable, and have demonstrated its superior performance under a wide range of conditions. The stability result is remarkably improved from the behavior of other RPE algorithms, especially with poles close to the unit circle and, in particular, in comparison with previous adaptive notch filters. These advantages are due to the special constraints imposed on the filter and the use of a proper as well as a minimal model. These results raise the question of whether some larger subgroup of the RPE algorithms does not require stability monitoring. The success of the model used in this paper motivates the extension to other constraints and more general forms. These subjects are currently being investigated.

APPENDIX

In this Appendix, we derive the CRB results (20) and (22) for the narrow-band ARMA process (18).

For ARMA processes, it can be shown (see, for example, [6]) that

$$J = NE[\psi(t) \psi^T(t)] \quad (A1)$$

where $\psi(t) = -\partial e(t)/\partial \theta$. This expression can be obtained by noting that the cost function (3b) is essentially equivalent to the negative log-likelihood function of the ARMA signal. In our case, the results (10) and (11) of the gradient vector are applicable for the special signals in the form of (16), except that $y(t-i)$ and $\epsilon(t)$ have to be replaced by $r^i y(t-i)$ and $e(t)$, respectively.

From the above and using (10), one obtains for the process (18)

$$\begin{aligned} \psi(t) &= \frac{\rho e(t-1) - ry(t-1)}{A(\rho q^{-1})} \\ &= \frac{(\rho - r)(1 - \rho r q^{-1})}{A(\rho q^{-1}) A(r q^{-1})} e(t-1) \\ &= \frac{B(q^{-1})}{C(q^{-1})} e(t-1) \end{aligned} \quad (A2)$$

where $C(q^{-1}) = A(\rho q^{-1}) A(r q^{-1})$ and $B(q^{-1}) = (\rho - r)(1 - \rho r q^{-2})$. Notice that $\psi(t)$ has an ARMA form, while for general unconstrained ARMA signals, it would be an autoregressive (AR) signal (see, for example, [6]).

Now applying (A1) and (A2), we have

$$J = \frac{N}{2\pi j} \oint \frac{B(z^{-1}) B(z)}{C(z^{-1}) C(z)} \frac{dz}{z} \quad (A3)$$

where \oint denotes integration on the unit circle. Expression (A3) can be written alternatively

$$J = N b^T R b \quad (A4a)$$

where

$$b = (\rho - r) [1 \ 0 \ -\rho r \ 0 \ 0]^T \quad (A4b)$$

and R is the covariance matrix of an AR process characterized by the denominator polynomial $C(q^{-1})$. The matrix R can be directly computed in terms of the coefficients of $C(q^{-1})$ using the useful formula derived in [11]:

$$R = [T_1 T_1^T - T_2 T_2^T]^{-1} \quad (A5)$$

where here T_1 and T_2 are 4×4 lower triangular Toeplitz matrices characterized by their first columns:

$$T_{11} = [1 \ c_1 \ c_2 \ c_3 \ c_4]^T \quad (A6a)$$

$$T_{21} = [0 \ c_4 \ c_3 \ c_2 \ c_1]^T \quad (A6b)$$

and c_i are the coefficients of $C(q^{-1})$. Thus, we can write the following close formula for the asymptotic error variance of \hat{a}_1 :

$$\text{var}(\hat{a}_1) = \frac{1}{N b^T [T_1 T_1^T - T_2 T_2^T]^{-1} b} \quad (A7)$$

In addition, we can compute the expected variance on the central frequency estimate \hat{f}_1 (i.e., the notch frequency) of our algorithm as follows. Factoring the polynomials in (18) gives

$$f_1 = \frac{1}{2\pi} \cos^{-1}(a_1/2). \quad (A8a)$$

Hence

$$\begin{aligned} \frac{\partial f_1}{\partial a_1} &= -\frac{1}{2\pi a_1 \tan \omega_1} \\ &= -\frac{1}{2\pi \sqrt{4 - a_1^2}} \end{aligned} \quad (A8b)$$

Using linearization of f_1 about a_1 , the asymptotic variance of \hat{f}_1 is given by

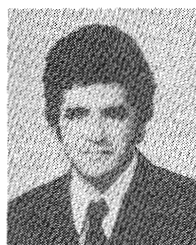
$$\begin{aligned} \text{var}(\hat{f}_1) &= \left[\frac{\partial f_1}{\partial a_1} \right]^2 \frac{1}{J} \\ &= \frac{1}{4\pi^2 N (4 - a_1^2) b^T [T_1 T_1^T - T_2 T_2^T]^{-1} b} \end{aligned} \quad (A9)$$

ACKNOWLEDGMENT

The author wishes to acknowledge useful discussions with J. O. Smith on the subject of this paper.

REFERENCES

- [1] P. A. Thompson, "A constrained recursive adaptive filter for enhancement of narrow-band signals in white noise," in *Proc. 12th Asilomar Conf. Circuits, Syst. Comput.*, Pacific Grove, CA, Nov. 1978, pp. 214-218; also Ph.D. dissertation, Elec. Eng. Dep., Stanford University, CA, February 1979.
- [2] S. Y. Kung and D. V. Bhaskar Rao, "An unbiased adaptive method for retrieval of sinusoidal signals in colored noise," in *Proc. 20th IEEE Conf. Dec. Control*, San Diego, CA, pp. 801-807; also in *Proc. 1982 IEEE Int. Conf. Acoust., Speech, Signal Processing*, Paris, May 1982, pp. 663-666.
- [3] D. V. Bhaskar Rao and S. Y. Kung, "Adaptive notch filtering for the retrieval of sinusoids in noise," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-32, pp. 791-802, Aug. 1984.
- [4] B. Friedlander and J. O. Smith, "Analysis and performance evaluation of an adaptive notch filter," *IEEE Trans. Inform. Theory*, vol. IT-30, pp. 283-295, Mar. 1984.
- [5] L. Ljung, "Analysis of a general recursive prediction error identification algorithm," *Automatica*, vol. 17, no. 1, pp. 89-100, Jan. 1981.
- [6] L. Ljung and T. Söderström, *Theory and Practice of Recursive Identification*. Cambridge, MA, MIT Press, 1983.
- [7] R. Rabiner and B. Gold, *Theory and Application of Digital Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1975.
- [8] T. Söderström, L. Ljung, and I. Gustavsson, "A comparative study of recursive identification methods," Dep. Automatic Control, Lund Inst. Tech., Lund, Sweden, Rep. 7427, 1974.
- [9] A. Nehorai, "Algorithms for system identification and source location," Ph.D. dissertation, Elec. Eng. Dep., Stanford University, CA, June 1983.
- [10] L. Ljung, "Analysis of recursive stochastic algorithms," *IEEE Trans. Automat. Contr.*, vol. AC-22, pp. 551-575, Aug. 1977.
- [11] T. Kailath, A. Viera, and M. Morf, "Inverses of Toeplitz operators, innovations, and orthogonal polynomials," *SIAM Rev.*, vol. 20, pp. 106-110, Jan. 1978.
- [12] D. C. Rife and R. R. Boorstyn, "Multiple tone parameter estimation from discrete-time observations," *Bell Syst. Tech. J.*, vol. 55, no. 9, pp. 1389-1410, Nov. 1976.
- [13] B. Widrow *et al.*, "Adaptive noise cancelling: Principles and applications," *Proc. IEEE*, vol. 63, pp. 1692-1716, Dec. 1975.
- [14] A. Nehorai and D. Malah, "On the stability and performance of the adaptive line enhancer," in *Proc. 1980 IEEE Int. Conf. Acoust., Speech, Signal Processing*, Denver, CO, Apr. 1980, pp. 478-481.
- [15] S. M. Kay and S. L. Marple, Jr., "Spectrum analysis—A modern perspective," *Proc. IEEE*, vol. 69, pp. 1380-1419, Nov. 1981.
- [16] P. Stoica and T. Söderström, "On the parsimony principle," *Int. J. Control*, vol. 36, no. 3, pp. 408-418, 1982.



Arye Nehorai (S'80-M'83) was born in Haifa, Israel, on September 10, 1951. He received the B.Sc. and M.Sc. degrees in electrical engineering from the Technion-Israel Institute of Technology, Haifa, in 1976 and 1979, respectively, and the Ph.D. degree in electrical engineering from Stanford University, Stanford, CA, in 1983.

From 1976 to 1979, he was a Teaching Assistant with the Department of Electrical Engineering, Technion. From 1979 to June 1983, he was a Research Assistant, and from June 1983 to January 1984, he was a Research Associate at Stanford University. He was responsible for the development of new algorithms for system identification, adaptive filtering, multisource location, and analysis of fast estimation algorithms. In January 1984, he joined Systems Control Technology, Inc., in Palo Alto, CA, working in performance analysis of missile systems. His areas of interest are statistical signal processing, digital filtering, and system identification.

Dr. Nehorai is a recipient of the Rothschild and Willer Fellowships and the Most Outstanding Graduate Student Award in the Department of Electrical Engineering, Technion. He is a member of Sigma Xi.