ANALYSIS AND IMPLEMENTATION OF THE ADAPTIVE NOTCH FILTER FOR FREQUENCY ESTIMATION*

Sun-Yuan Kung and D.V. Bhaskar Rao

Department of Electrical Engineering-Systems
University of Southern California
Los Angeles, California 90007

ABSTRACT

This paper enhances some theoretical and implementation aspects of a constrained autoregressive moving average model, the notch filter model developed in [1] for the estimation of sinusoidal signals in additive, uncorrelated noise, colored or white. This model is shown to approximate the actual signal plus noise model. In addition, the parameter estimates obtained by minimization of the output power of the notch filter approximate the maximum likelihood estimate of the model parameters. The relationship of the notch filtering approach to the existing autoregressive and Pisarenko methods is established. Next, a scheme to combine fast convergence and unbiased estimation is suggested. Lastly, certain implementation aspects of the filter are considered and the method is shown to be amenable to parallel processing.

INTRODUCTION

Many new techniques have been developed in the last two decades for spectral analysis of discrete time series. Parametric methods, in particular where an autoregressive (AR) model is assumed, have been shown to exhibit high resolution and high performance with short data length. For adaptive AR parameter estimation, probably the most familiar structures are the tapped delay line or transversal filter (e.g. [2]), and the more recent lattice filtering techniques based on the Burg's maximum entropy method [3]. They all assume an underlying AR model and are closely related to the linear prediction techniques in speech.

In a noisy environment, the AR based methods mentioned above, lead to biased estimates and lower resolution [4] [5]. Various methods have been suggested to compensate for this effect [6] [7] [8]. There are, however, limitations to such approaches. The estimates are biased because the all pole model is not appropriate: the more compatible choice is a pole-zero model, or autoregressive moving-average (ARMA) model. Unfortunately, though ARMA methods are the most desirable, they are far more complex.

Success has been achieved by restricting applications to a special class of signals. This paper focusses on a particular and yet a very important class of spectrum analysis problems: the estimation of the frequencies of sinusoidal signals in additive noise, colored or white. In fact for the white noise case, Pisarenko's harmonic decomposition has been gaining popularity, for unbiased estimation of the sinusoidal frequencies [9] [10]. However, it is sensitive to the estimates of the autocorrelation [11], and a data-adaptive technique is desirable to obtain the solution recursively. In this case, a filter based on a special ARMA model is of interest. This paper considers such a special ARMA model, a constrained ARMA model, namely a notch filter model introduced in some earlier publications [1],[12]. The purpose of the present paper is to further explore both theoretical and implementation aspects of the notch filtering technique.

THEORETICAL BASIS FOR THE NOTCH FILTER

Frequency Domain Analysis

The notch filter model developed in [1] is briefly reviewed. 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 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. Since the noise is uncorrelated with the signal,

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

Also, $S_x(\omega)$ which represents the sinusoidal signals is given by

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

On passing the signal y_t through a digital filter $H(\phi)$, ϕ being the parameter vector governing the filter, the output power J is given by

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

Using Eq.(2) and Eq.(3) we have

$$\begin{aligned} &= \tfrac{1}{2\pi}\int S_n(\omega)|H(e^{j\omega})|^2 d\omega + \sum_{i=1}^{P} \frac{P_i}{2\pi} |H(e^{j\omega_i})|^2 \\ &= A(\phi) + B(\phi) \end{aligned} \tag{4}$$

where $A(\phi) = \tfrac{1}{2\pi}\int S_n(\omega)|H(e^{j\omega})|^2 d\omega$,

$B(\phi) = \sum \frac{P_i}{2\pi} |H(e^{j\omega_i})|^2$

Unless special precautions are taken, minimization of J will result in biased estimates of the sinusoidal frequencies, e.g. bias due to AR modeling [1]. If the parameter vector is constrained to create a (extemely sharp) notch

*This research has been supported in part by the Office of Naval Research under Contract No. N0014-81-K-0191; ONR Contract No. N0014-80-C-0457; and by Army Research Office under Grant No. DAAG 29-79-C-0054.

filter tranfer-function (response) i.e.

$$H(e^{j\omega}) = \begin{cases} 0, & = \omega'_1, \ldots, \omega'_p \\ 1, & \text{elsewhere} \end{cases} \qquad (5)$$

then the effect of noise can be eliminated. On substituting Eq.(5) in Eq.(4), $A(\Phi)$ now becomes a constant, independent of Φ (and equals the noise power). The output power J is given by $J = A + B(\Phi)$

and $\text{Min } J = A + \text{Min } B(\Phi) = A$

The minimizing condition is that the transfer function $H(e^{j\omega})$ have it zeros at frequencies $\omega_1, \ldots, \omega_p$ making $B(\Phi) = 0$ (=minimum). This implies that minimization of J does ensure an unbiased estimate of the 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.

Realization

The ideal notch filter is impractical, so an approximate realization is suggested [1]. The transfer function of the realization is given by

$$H(z^{-1}) = (1 - W(z^{-1}))/(1 - W'(z^{-1})) \qquad (6)$$

where

$$1 - W(z^{-1}) = 1 - w_1 z^{-1} - w_2 z^{-2}, \ldots - w_p z^{-p} = \prod_{i=1}^{p} (1 - e^{j\omega_i} z^{-1})]$$

$$1 - W'(z^{-1}) = 1 - \alpha w_1 z^{-1} - \alpha^2 w_2 z^{-2} \ldots \alpha^p w_p z^{-p} = \prod_{i=1}^{p} (1 - \alpha e^{j\omega_i} z^{-1}), \quad 0 \leq \alpha < 1$$

defining

$W = (w_1, w_2, \ldots, w_p)$, feedforward coefficients

$W' = (w'_1, w'_2, \ldots, w'_p)$, feedback coefficients.

The feedforward and feedback coefficients are related

$$w'_i = \alpha^i w_i, \quad 0 \leq \alpha < 1$$

This constraint on the coefficients restricts the zeros and poles to lie on the same radial line, their separation being governed by α. For stability reasons α should be less than one. Qualitatively, one can expect that as $\alpha \rightarrow 1$, the notch filter is sharper and is a better approximation, resulting in estimates that are less biased. As α gets closer to zero, the noise power has greater effect. Therefore, α is called the debiasing parameter.

Comparison with the Time Domain Model

A comparison of the notch filter model to the actual signal plus noise model is useful in understanding the nature of the approximation involved. Assuming there are exactly P/2 sinusoids

$$x_t = W_a^T X_t$$

where

$W_a = (w_{a1}, w_{a2}, \ldots, w_{ap})^T$ is the actual weight vector, and

$X_t = (x_{t-1}, x_{t-2}, \ldots, x_{t-p})^T$ is a vector containing p lag values

Y_t and N_t are similarly defined. Eq.(1) then becomes

$$y_t = W_a^T X_t + n_t = W_a^T Y_t - W_a^T N_t + n_t$$

So the difference equation of the actual model is

$$n_t = y_t - W_a^T Y_t + W_a^T N_t \qquad (7)$$

On the other hand, from eq. (6) it can be seen that the difference equation governing the output of the notch filter $\hat{n}_t$, i.e. the notch model, is given by

$$\hat{n}_t = y_t - W^T Y_t + W^T \Lambda \hat{N}_t \qquad (8)$$

where $\Lambda = \begin{pmatrix} \alpha & & 0 \\ & \alpha^2 & \\ 0 & & \ddots \alpha^p \end{pmatrix}$

A comparison of Eq.'s(7) and (8) show that as $\alpha \rightarrow 1$, the difference equation governing the notch filter approaches that of the actual model.

Maximum Likehood Estimates of W

The difference equation governing the given signal y_t is a special ARMA (p,p) process given in (7). Like in any other parameter estimation problem, a maximum likelihood estimate of $\underline{W}$ can be derived when $\{n_t\}$ is a sequence of independent normal $(0, \sigma^2)$ random variables, as in [13], [14]. Given a particular set of data Y_ℓ, the log likelihood function associated with the parameters (W,), conditioned on the choice of starting values $(\bar{Y}_*, N_*)$, is

$$L(W, \sigma) = -\ell \ln\sigma - \sum_{t=1}^{\ell} n^2_t(\underline{W}, \sigma | Y_*, N_*, Y_\ell)/2\sigma^2$$

The point that requires emphasis is that the cost function to be minimized

$$J = \sum_{t=1}^{\ell} n_t^2(\underline{W}, \sigma | Y_*, N_*, Y_1)$$

depends very critically on the starting values (Y_*, N_*). For most purposes, the dependence on the starting values (Y_*, N_*) can be avoided by using a reasonable value (unconditional expectations). However, in this particular case, as the poles are on the unit circle, the starting values are of extreme importance, as improper starting values could introduce large transients [13]. This requirement is hard to meet. Furthermore, the resulting filter is unstable and the method requires an exhaustive search using all values of $\underline{W}$, (Similar to the method in [15]). The notch filter can be perceived as an approximation to the ML estimator suggested by this parametric approach, so as to circumvent the need for an exhaustive search. Qualitatively, the closer α is to one $(\alpha \rightarrow 1)$, the closer the estimates approach the ML estimates, since the difference equation governing the notch filter approaches the actual model. Another interesting feature is that, constraining $\underline{W}$ to take values such that roots of $W(z)$ lie on the unit circle is equivalent to performing a maximum aposteriori estimate: the apriori probability density function of $\underline{W}$, being zero for all values of $\underline{W}$ which give rise to zeros inside or outside the unit circle and uniform, for all values of $W(z)$ with zeros on the unit circle. However, in the case of inaccurate apriori information regarding the number of sinusoids, determination of the amplitudes of the sinusoids becomes necessary to avoid false alarm.

Relation to Existing Methods

Interestingly, because of the debiasing parameter α, the notch model can be viewed as a

generalization of some of the popular methods.
Case I: $\alpha = 0$, the transfer function is given by

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

The structure is that of the familiar linear prediction filter.
Case II: $\alpha \to 1$, Under the assumption that the filter has converged to the desired solution, an approximation of the following nature can be made. $\hat{n}_t = n_t$ and $\Lambda = I$, Rearranging Eq.(8) we have

$$(y_t, Y_t)\begin{pmatrix}1\\w\end{pmatrix} = (n_t, N_t)\begin{pmatrix}1\\w\end{pmatrix} \tag{9}$$

Multiplying both sides of (9) by $\begin{pmatrix}y_t\\Y_t\end{pmatrix}$, taking expectations and using the uncorrelatedness of n_t and y_t, we have

$$R_y\begin{pmatrix}1\\w\end{pmatrix} = R_N\begin{pmatrix}1\\w\end{pmatrix} \text{or} (R_y - R_N)\begin{pmatrix}1\\w\end{pmatrix} = 0 \tag{10}$$

where $R_y = E[\underset{Y_t}{y_t}\,(y_t, Y_t)]$, $R_N = E[\underset{N_t}{n_t}\,(n_t, N_t)]$

For the white noise case, $R_N = \sigma^2 I$ and Eq.(10) reduces to $(R_y - \sigma^2 I)\begin{pmatrix}1\\w\end{pmatrix} = 0$.

This is similar to the Pisarenko harmonic retrieval method. From equation (10), as $\alpha \to 1$, the notch filter can be viewed as a generalized version of Pisarenko's method.

Convergence Rate

A gradient technique has been suggested for adaptively adjusting the coefficients of the notch filter [1]. A zero-pole (direct form) realization is chosen for the implementation of the notch filter. The convergence rate, for a fixed values of μ, is found to depend on the debiasing parameter α. An analysis of the dependence suggests a mechanism to combine fast convergence and unbiased estimation. From equation (8)

$$J = E(\hat{n}_t^2) = E[(y_t - \underline{W}^T Y_t + \underline{W}^T \Lambda \hat{N}_t)^2]$$
$$= \underline{W}^T A \underline{W} + \underline{W}^T \underline{B} + C.$$

The cost function can be written in a quadratic form as shown above. The gradient update can be summarized as follows

$$W_{t+1} = W_t - \mu \nabla J = W_t - \mu (2\, A\, W_t)$$
$$= (I - 2\mu A) W_t \tag{11}$$

It is well known that the step size μ should lie in the range: $0<\mu<1/\lambda_{max}$, where λ_{max} is the maximum eigenvalue of A. However, from (11), it can be seen that the convergence rate depends on the slowest "mode" of $I - 2\mu A$, i.e. $1 - 2\mu\lambda_{min}$, However, here A and B are not constants but are function of $\underline{W}$ which make the analysis difficult.
$\alpha = 0$: This is the common linear prediction approach. Here $A = E(Y_t, Y_t^T) = R_y$ and so convergence rate depends on λ_{min} of R_y.
$\alpha \to 1$: Here the analysis is divided into two parts.
Case I - $\underline{W}$ far from the optimal solution. Because the notch filter is a narrowband rejection filter, $\hat{n}_t$ can be approximated by $\hat{n}_t = y_t$ and $\Lambda = I$. With this approximation, "A" becomes nearly singular making λ_{min} small and the convergence very slow.
Case II: $\underline{W}$ close to the optimal solution. Here, approximation $\hat{n}_t = n_t$ is made resulting in $A = R_y - R_N$ which gives rise to a λ_{min} close to $\alpha=0$ condition.

Now, from the above examination of λ_{min}, the following inference can be made. For $\alpha \to 1$ and fixed step size given by the bounds of Case II, it can be seen that the initial convergence will be slow (from Case I) and acceptable in the latter stage(i.e. near the optimal solution from Case II). Since the convergence for $\alpha = 0$ is faster in general, it is appropriate to start with $\alpha = 0$ and then increase α as the estimate is approached, to get an improved convergence rate. Furthermore, by using a recursive least square type of approach, faster convergence can be achieved at the expense of a more complex updating mechanism.

A heuristic explanation to this is as follows: when $\alpha \to 1$ and $\underline{W}$ is far from optimal, being a narrow band rejection filter it is unable to sense the sinusoidal signals because of its flat response outside the notch frequencies. On the other hand $\alpha = 0$ implies a larger bandwidth and so is capable of sensing the sinusoids. However, α close to 1 and $\underline{W}$ near optimal, the sinusoids are within the narrowbands of the notch filter and are sensed adequately. So a starting value $\alpha = 0$ is very convenient. Though the above derivation and explanation are not rigorous, they provide insight into the understanding of the filter's performance.

IMPLEMENTATION

The notch filter being IIR (infinite impulse response) in nature, and in particular, the poles and zero's being close to the unit circle, its realization requires special attention. Various aspects of filter design, dynamic range, stability etc. have to be taken into account.

Direct Form

Due to dynamic range considerations, a zero-pole (direct form) realization is preferred, particularly when $\alpha \to 1$, for the implementation of the notch filter [1]. The above realization performs adequately for single sinusoids or multiple sinusoids whose frequencies are fairly separate. For closely spaced sinusoids, parametric sensitivity of direct forms renders the filter unstable [1]. To circumvent the parametric sensitivity problem, a smaller value of α is preferable. However a small value of α does reduce the resolution and accuracy. Intuitively a larger length filter should help compensate this drawback. For small values of α, a pole-zero configuration is equally well suited. In fact, the gradient generating mechanism can be simplified [16]. For high resolution (closely spaced sinusoids and $\alpha \to 1$), parallel and cascade forms are very suitable. A study of the behavior of the notch filter, when undermodelled, is useful for such implementations.

Local Minima

For a signal consisting of p sinusoids, if a second order filter is used i.e. a notch filter capable of eliminating one sinusoid is used, then there exist many local minima. Using a frequency domain analysis as before, it is not hard to see that each minimum corresponds to the removal of one of the sinusoids. In a practical case, it holds only when α is close to one.

Parallel Forms

In addition to avoiding the problem of parametric sensitivity, the parallel realization of second order filters as in fig.(1) offers several advantages. The local minima property, discussed above, ensures that each filter removes one sinusoid. The local minimum each section converges to depends on the starting point. Therefore, the entire frequency domain can be divided into small zones. Each zone is monitored by one second order notch filter and restricted to scan a particular frequency region by a constraint on the coefficients of the filter. This parallel processing scheme bears a strong resemblance to the adaptive frequency sampling filters in [17]. The output power of each section can be independently reduced because of the frequency decoupling property i.e. none or negligible overlap in the bandwidth of the filters.

Cascade Forms

The same local minima property can be exploited for a cascade configuration. A Cascade form as indicated in Fig.(2) can be useful. The first section cancels one sinusoid and transmits the rest of the signal with little distortion, to the second section which cancels a second sinusoid. Again, each section can be independently minimized. The output of section I can be used to adaptively adjust its coefficients, while the output of section II can be used to determine the filter coefficients of section II.

However, a large cascade realization is not feasible in practice because the approximation becomes less valid as more sections are added. So a cascade network in conjunction with the parallel form may be useful. The parallel form is useful for scanning the entire frequency spectrum, while a cascade form can be used for each frequency bin to resolve closely spaced sinusoids

To summarize the implementation aspects, a zero-pole implementation requires less hardware and is suitable for operations where signals are well separated. The possibility of adopting a intermediate value of α to obtain partial debiasing appears feasible, although coupling between coefficients does exist. On the other hand, the parallel and cascade forms require $\alpha \to 1$, which makes the minimization process complex [1]. It also results in more hardware because frequency zones have to be scanned irrespective of whether a signal is present or not in that zone. However, each section can be independently minimized as the above mentioned coupling is removed. Furthermore, with VLSI and for real time processing, it may be a viable candidate.

SIMULATION

Some encouraging results have been observed in the simulations conducted. They clearly demonstrate the debiasing effect of α and also its effect on the convergence rate. The parallel and cascade forms have also appeared to be promising, particularly for closely spaced sinusoids. Due to lack of space, the simulation results will be presented at the conference. For some earlier results, the reader is referred to [1].

REFERENCES

1 Kung, S.Y and D.V.Bhaskar Rao, pp 801-806, Proc., IEEE 20th CDC, San Diego, CA, Dec. 16-18, 1981.

2 Griffiths, L.J., pp. 207-222, IEEE Vol. ASSP-23, April 1975.

3 Burg, J.P. " Maximum Entropy Spectral Analysis ", Ph.D. Dissertation, Dept. of Geophysics, Stanford Univ., Stanford, CA, May 1975.

4 Pagano, M., pp. 99-108, Ann. Statistics, Vol. 2, 1974.

5 Kay, S.M., pp. 478-485 IEEE Trans. Vol. ASSP-27, October 1979.

6 Kay, S.M., pp. 357-360 , in Rec. 1978 Int. ICASSP.

7 Treichler, in Rec. 1979 ICASSP, pp. 933-936.

8 Reddy, V.U., B. Egardt and T. Kailath, pp 702-709, IEEE Trans. Vol. ASSP-29, Part III, June 1981.

9 Pisarenko, V.F., Geophys. J. R. Astron. Soc., Vol. 28, pp. 511-531, 1972.

10 Pisarenko, V.F., Geophys. J. R. Aston. Soc. Vol. 33, pp. 347-366, 1973.

11 Ulrych, T.J. and R. W. Clayton, Phys. Earth Planet. Inter., Vol. 12, pp. 188-200, August 1976.

12 Kung, Sun-Yuan and D.V. Bhaskar Rao, pp 1-6, session 14, Vol. II, IFAC New Delhi, India, January 5-7, 1982.

13 Box, G.E.P. and Jenkins, G.M., " Time Series Analysis: Forecasting and Control ", San Francisco, CA; Holden-Day, 1970.

14 Astrom, K.J. and T. Bohlin, Proc. 2nd IFAC Symposium; 1965 NPL Tedd England, Plenum Press; New York; pp. 96-111.

15 Viterbi, A.J., " Principles of Coherent Communication ", New York: McGraw-Hill; 1966.

16 Hsia, T.C., pp. 1153-1155, letters to Proc. IEEE; Vol. 69, No. 9; Sept. 1981.

17 Bitmead, R.R and B.D.O Anderson, pp. 684-694, Trans. Vol. ASSP-29, June 1981.

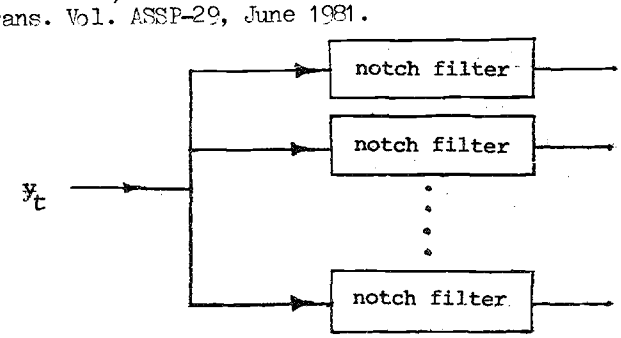

Fig. 1: Parallel form of second-order notch filters

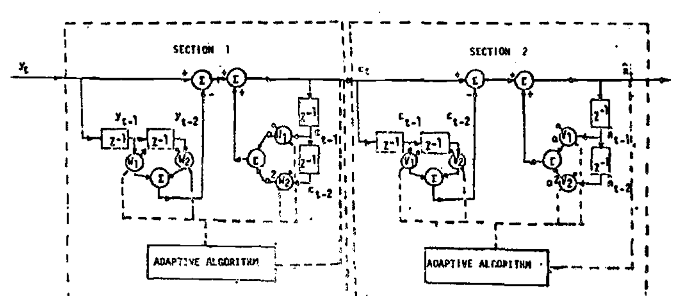

Fig. 2: Cascade Realization