(1)

FINAL REPORT

STUDIES OF ADAPTIVE SYSTEMS

University of Rochester

Principle Investigator: Prof. Charles W. Merriam
(716)275-4054

Project Engineer: Darell Marsh
(714)225-6301

ARPA Order Number: 1872
Program Code Number: 5G10
Contract Number: N00123-75-C-1282
Date of Contract: 5/16/75
Contract Expiration Date: 3/31/77
Amount of Contract: $42,240

Sponsored by

Advanced Research Projects Agency
ARPA Order No. 1872

The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the Advanced Research Projects Agency or the U.S. Government.

This research was supported by the Advanced Research Projects Agency of the Department of Defense and was monitored by the Naval Ocean Systems Center under Contract. N00123-75-C-1282.

Overview

This report is organized as five separate reports covering the five tasks of the contract. Four of these reports are included in the present document plus a separate report written by Prof. C. W. Merriam, III.

The first section of this report provides the asymptotic theory and geometric procedure for determining the amplitude of the crosscorrelation function for FM signals whose instantaneous frequency curves overlap at one point in time. In addition, the geometric procedure can be used to upperbound the cross-correlation function for multiple intersecting cases, where absolute phase of the signal components is not generally known.

The second section provides a new definition of wideband energy density function which allows for Doppler stretch. The key to this definition and its properties lies in the use of the Fourier-Mellon transform which converts time stretching to a multiplicative exponential factor in the transform domain. The relationship between the wideband energy density function and the wideband ambiguity function is also established using a modified two-dimensional Fourier transform. Using these new results, we feel we can now extend the theory of wideband signal processing.

The third section discusses autoregressive techniques including the maximum entropy predictive technique. Although no new analytic results were found, the extensive simulations show that these techniques have certain difficulties, particularly in resolving closely spaced echos in a multipath environment. Thus,

except for particular circumstances, these techniques do not do as well as a DFT. Caution should be used in applying maximum entropy estimation to periodicity estimation. Included in this section is a complete review of the multiple ways of viewing the overall problem of autoregressive spectral analysis.

The fourth section has several new results concerning adaptive estimation procedures. An exponentially weighted least squares algorithm (EWLS) is developed. The Widrow-Hoff algorithm (LMS) is an approximation to the EWLS algorithm. Several important results are presented. In particular, the importance of choosing an input signal sequence whose wideband ambiguity function is insensitive to Doppler scaling is established. Other results include "misadjustment noise", stability, error in tracking time-varying parameters, effects of additive noise, and the role of the input signal.

The fifth section, included as a separate report, covers other gradient algorithms in parameter estaimtion. Included here is a discussion of the effects of coefficient averaging and the reduction of the misadjustment effects.

The only area in which we feel that definitive results did not occur is in determining the effects of Doppler stretch on the Wiener solution. The difficulty here is that although the original signal and the Doppler scaled signal are stationary, they are not jointly stationary. Thus, all efforts at formulating the problem analytically failed, despite the fact the much initial effort was devoted to this problem.

Finally, the authors would like to acknowledge the help of several people without whose help this research could not have been done. First, the help of the staff of the Perinatal computing facility is acknowledged. Equipment funds from this contract were used to purchase several peripheral pieces of equipment for this facility and can now be used for future Navy contracts. We would like to thank Dr. R. A. Altes for providing his manuscript covering the Fourier-Mellon transform which proved to be the key to the determination of the wideband energy density function. Finally, we wish to thank Darrell Marsh for his patience in allowing the authors to complete this work during a period of difficult time for us.

Also, we would like to give a special thanks to Beth Dunn for her typing of this report.

Asymptotic and Geometric Procedures for Estimating the Crosscorrelation Function of Frequency Modulated Signals

by

Edward L. Titlebaum
Associate Professor of Electrical Engineering
University of Rochester
Rochester, NY 14627

I. Introduction

For many Navy applications, Frequency Modulated (FM) signals which have large time-bandwidth product are being considered for underwater acoustic communications. With the possibility of coherent processing and a large number of signals, it is clearly desirable to find ways of evaluating the mutual crosscorrelation functions of various signals. Using the method of stationary phase we will derive an asymptotic formula for the crosscorrelation of two FM signals whose instantaneous frequency curves cross at only one point in time. For this case, a simple geometric procedure for estimating the crosscorrelation is given based upon the overlapping area of two templates generated in time-frequency space.

For signals with more than one instantaneous frequency crossing, the geometric procedure provides an upperbound on the crosscorrelation, since, in general the absolute phase between the two signals may not be known.

Since the cross ambiguity function is a crosscorrelations function we can apply this procedure directly to this case.

II. Asymptotic Theory.

In this section we consider the problem of asymptotically estimating the crosscorrelation function between two FM signals, based upon crossings of their instantaneous frequency curves. We shall deal with complex (analytic) signals. In Appendix A we establish the relationship between the crosscorrelation of the real parts of the analytic signals and the results for analytic signals themselves.

We begin by defining instantaneous frequency. Suppose we have an analytic signal of the form

$$f(t) = a(t)e^{j\theta(t)} \tag{1}$$

By analytic we mean that the Fourier transform of $f(t)$, $F(\omega)$, satisfies the condition

$$F(\omega) = 0, \quad \omega < 0 \tag{2}$$

Thus the signal has a one-sided spectrum. The instantaneous frequency of $f(t)$ is the time-rate of change of its phase, thus

$$\omega_f(t) = \frac{d\,\theta(t)}{dt} \tag{3}$$

Thus, for example if

$$\theta(t) = \frac{B}{2}\,t^2 + wt \tag{4}$$

for a typical chirp FM signal, then

$$\omega_f(t) = Bt + w \tag{5}$$

We assume that the two signals of interest are of the same form:

$$\begin{cases} f(t) = a(t)e^{jA\alpha(t)} & (6\text{-a}) \\ g(t) = b(t)e^{-jB\beta(t)} & (6\text{-b}) \end{cases}$$

and the cross correlation is given by

$$R_{fg}(\tau) = \int_{-\infty}^{\infty} f(t)\, g^*(t+\tau)\, dt \tag{7}$$

Without loss of generality we assume the analytic signals have unit energy. Thus

$$\int_{-\infty}^{\infty} |f(t)|^2\, dt = \int_{-\infty}^{\infty} |g(t)|^2\, dt = 1 \tag{8}$$

Thus, by the Schwarz Inequality, the maximum value of their auto- and crosscorrelation functions is unity. Their instantaneous frequencies are thus

$$\omega_f(t) = A\,\frac{d\alpha(t)}{dt} \tag{9-a}$$

and

$$\omega_g(t) = -B\,\frac{d\beta(t)}{dt} \tag{9-b}$$

The two constants, A and B, are assumed to be large.

Substituting equations (6) into (7) yields

$$R_{fg}(\tau) = \int_{-\infty}^{\infty} k(t,\tau) e^{j\chi h(t,\tau)} dt \qquad (10)$$

where

$$\begin{cases} k(t,\tau) = a(t)b(t+\tau) & (11\text{-a}) \\ \chi h(t,\tau) = A\alpha(t) + B\beta(t+\tau) & (11\text{-b}) \end{cases}$$

If we assume that A>B, without loss of generality, then we have

$$\begin{cases} \chi = A & (12\text{-a}) \\ h(t,\tau) = \alpha(t) + \frac{B}{A}\beta(t) & (12\text{-b}) \end{cases}$$

Applying the method of stationary phase to the integral (4) we have, as an asymptotic expression

$$R_{fg}(\tau) \sim \left[\frac{2\pi}{\chi \frac{d^2 h(\hat{t},\tau)}{dt^2}}\right]^{1/2} k(\hat{t},\tau) \exp\left[j\chi h(\hat{t},\tau) + j\pi/4\right] \qquad (13)$$

where the stationary point $\hat{t}$ is the unique solution of the equation

$$\omega_f(\hat{t}) - \omega_g(\hat{t}) = 0 \qquad (14)$$

It is assumed that g is continuous and h is twice continuously differentiable with

$$\frac{d^2 h(\hat{t},\tau)}{dt^2} > 0$$

Let us do an example of how this procedure works. We will first look at the cross correlation between two linear FM signals, one an up chirp and the other a down chirp. We let

$$\begin{cases} f(t) = a(t)e^{j\frac{A}{2}t^2} \\ \\ g(t) = b(t)e^{-j\frac{A}{2}(T-t)^2}, \quad 0 \leq t \leq T \end{cases}$$

The instantaneous frequency curves for the two signals are shown in Figure 1.

The unit energy condition reveals that

$$\int_0^T a^2(t)dt = \int_0^T b^2(t)dt = 1$$

Solving for the stationary point, with χ=A we have that

$$h(t,\tau) = \frac{t^2}{2} + \frac{(t+\tau-T)^2}{2}$$

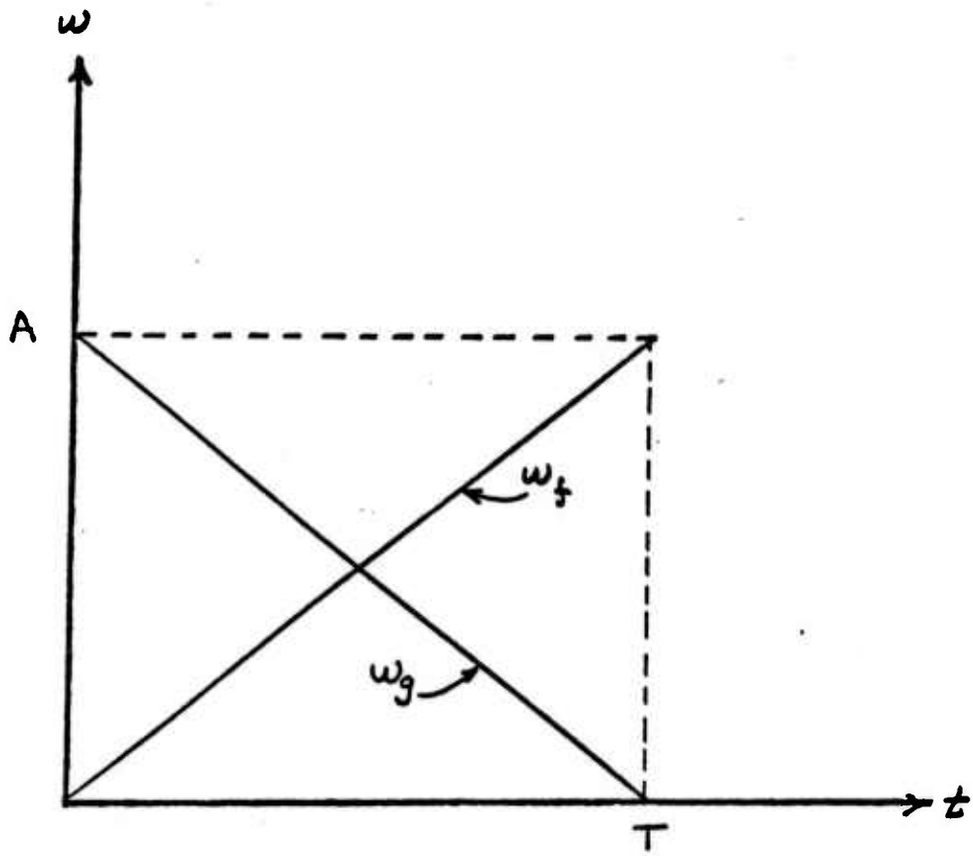
and differentiating

$$\frac{d\,h(t,\tau)}{dt} = t + (t+\tau-T)$$

Setting the derivative of h equal to zero and solving for $\hat{t}$ we obtain

$$\hat{t} = \frac{T-\tau}{2}$$

Figure 1. Instantaneous Frequency plots for Linear FM sweep example.



The second derivative condition is satisfied since

$$\frac{d^2h(t,\tau)}{dt^2} = 2 > 0$$

Solving for $k(\hat{t},\tau)$ and $h(\hat{t},\tau)$ we have

$$\begin{cases} k(\hat{t},\tau) = a(\frac{T-\tau}{2})\ b\ (\frac{T+\tau}{2}) \\ \chi h(\hat{t},\tau) = \frac{A}{4}\ (\tau-T)^2 \end{cases}$$

Thus we have as an estimate of R_{fg}

$$R_{fg}(\tau) \sim a(\frac{T-\tau}{2})\ b\ (\frac{T+\tau}{2})\ [\frac{\pi}{A}]^{1/2} \exp[j\frac{A}{4}(\tau-T)^2 + j\pi/4]$$

If we further assume that

$$a(t) = b(t) = \frac{1}{\sqrt{T}}, \quad 0 \leq t \leq T,$$

thus the signals have rectangular envelopes. Then in the region of overlap ($-T<\tau<T$) we have

$$|R_{fg}(\tau)| \sim \frac{1}{T}\ [\frac{\pi}{A}]^{1/2}$$

We observe that the total frequency deviation is AT rad/sec. or $2\pi F$ rad/sec.

$$|R_{fg}(\tau)| \sim \frac{1}{T}\ [\frac{\pi T}{2\pi F}]^{1/2} = \frac{1}{\sqrt{2FT}}$$

We see that, as the TB product gets large the crosscorrelation function diminishes as the reciprocal square-root of the TB product.

In order to check, the asymptotic results, several examples were run on our Interdata 7-32 minicomputer. The signals are of the form

$$F(I) = a(I)\exp[BI^2+WI],\ I = 1,\ N.$$

The amplitudes are rectangular except for a 10% raised cosine at either end of the signals. Figures 2. and 3. show the two signals for N=128. The real and imaginary parts are plotted consecutively on each graph. The parameters are

a) Upsweep B = 0.003, W = 0.3

b) Downsweep B = -0.003, W = 1.068

Calculating the timewidth and Bandwidth

$$T_f = (0.9)(128) = 115.2$$
$$F_f = (0.9)(\frac{0.768}{2\pi}) = 0.110$$

Thus the crosscorrelation estimate is

$$|R_{fg}| \sim \frac{1}{\sqrt{2T_fF_f}} = 0.1986$$

Figure 4. shows the actual crosscorrelation function and

the asymptotic estimate, shown as a straight line.

Suppose the 10% raised cosine window is not included then

$T_f = 128$

$F_f = 0.768$

and the estimate is $|R_{fg}| \sim 0.179$. This case is seen in figure 5., where the actual crosscorrelation and estimate are plotted together.

A second example with larger TB product is given below. Here the parameters are

N=256	a) Upsweep	B = .003, w = 0.300
	b) Downsweep	B = -0.003, w = 1.836

Figure 6. shows the first half of the crosscorrelation functions of the signals with and without the 10% raised cosine.

Figure 2. Real and Imaginary parts of 128 point Linear FM Upsweep signal.

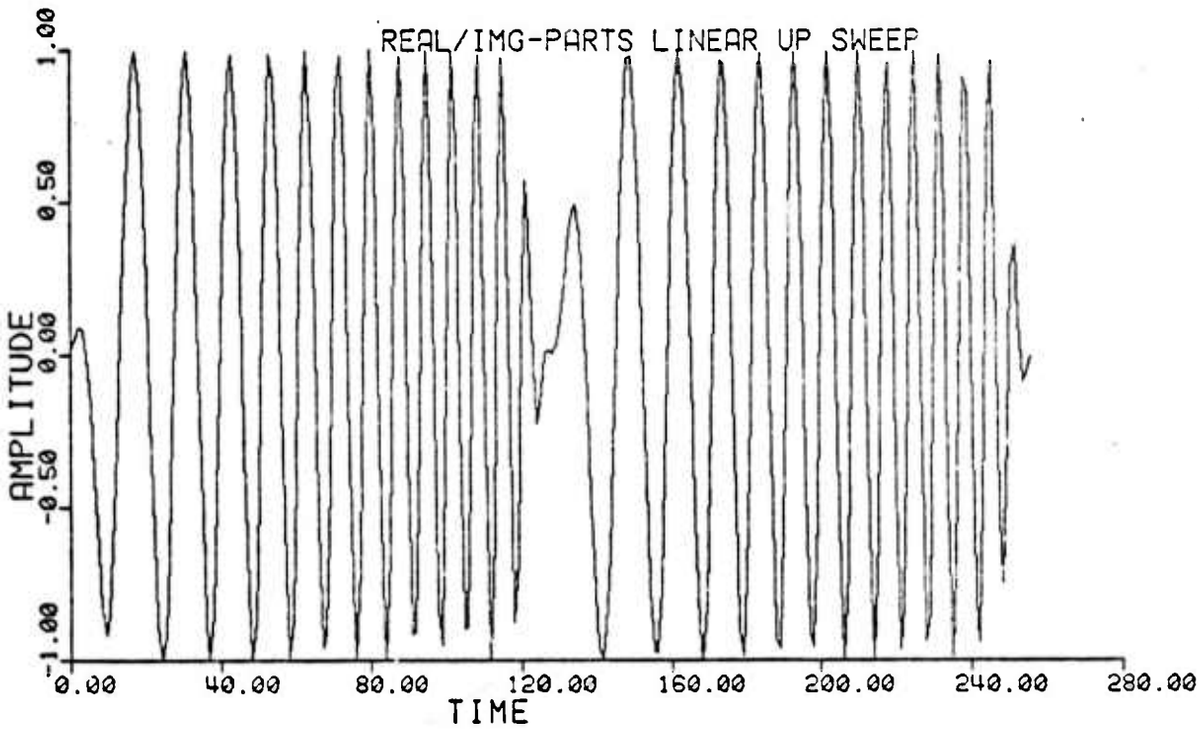


Figure 3. Real and Imaginary parts of 128 point Linear FM Downsweep signal

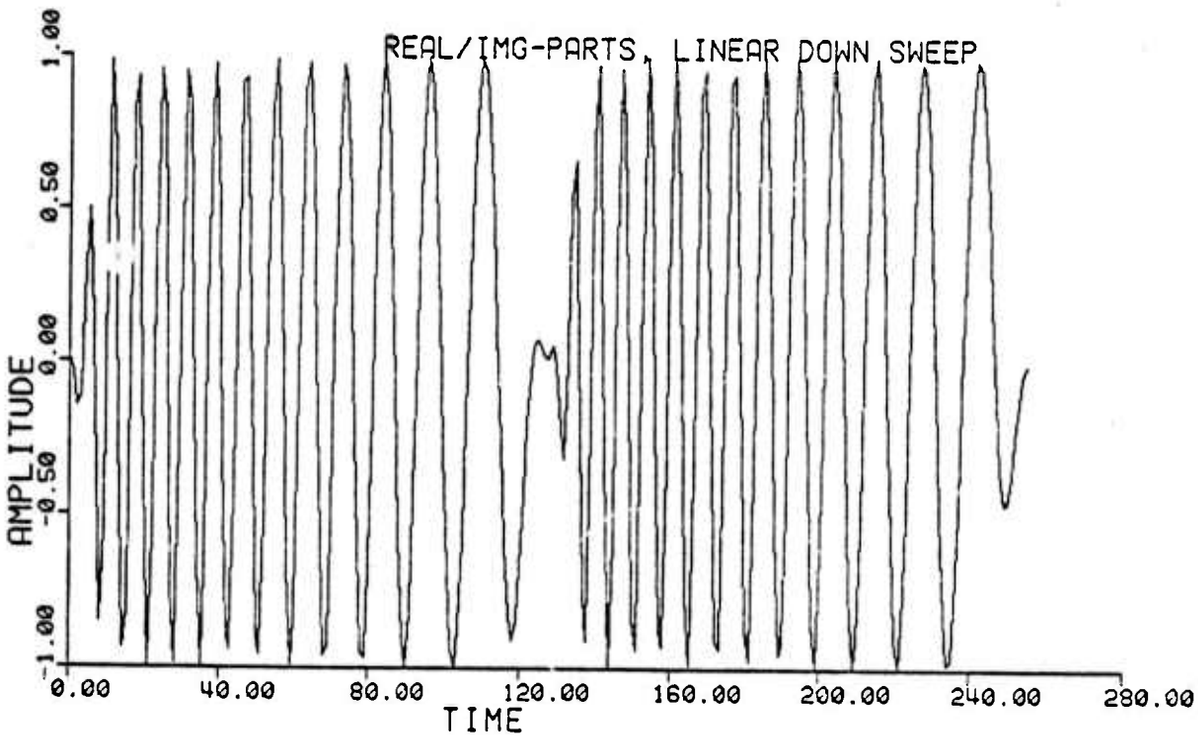


Figure 4. Crosscorrelation function and Asymptotic Estimate of Linear FM Up and Downsweep with 10% raised cosine window. $\tau=0$ corresponds to Time Delay = 127.

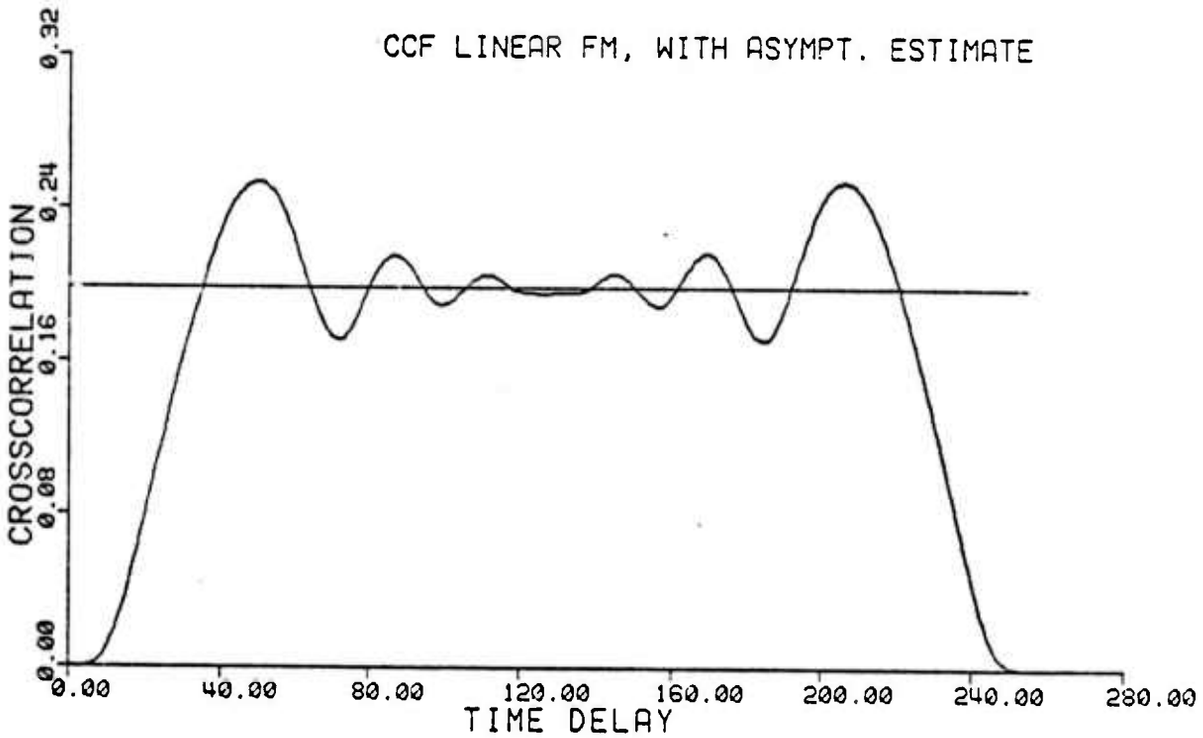


Figure 5. Crosscorrelation function and Asymptotic Estimate of Linear FM Up and Downsweep; NO window, τ=0 corresponds to Time Delay = 127.

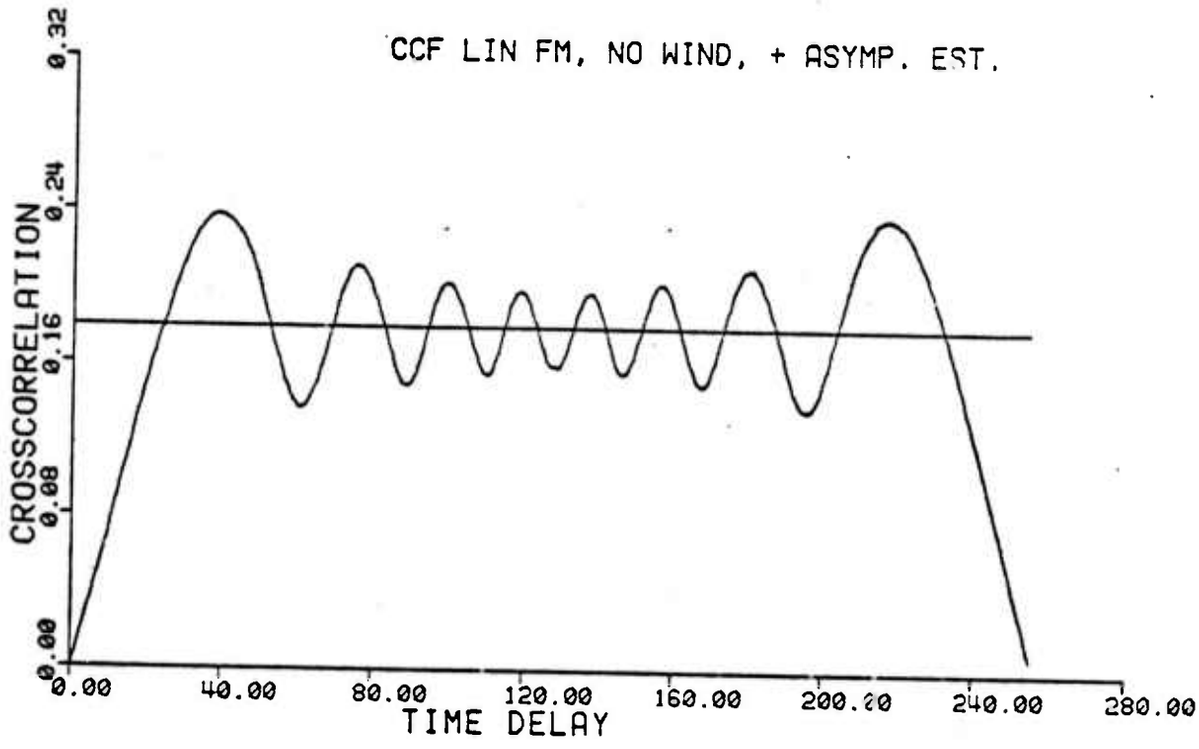
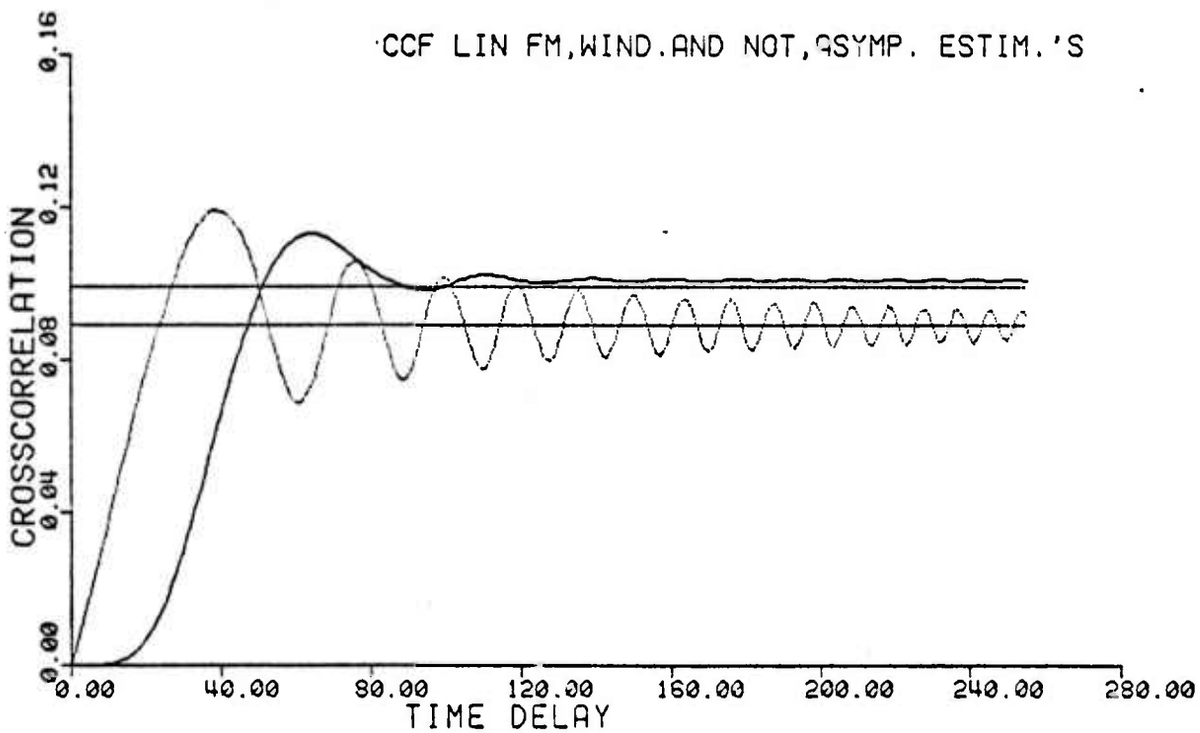


Figure 6. Half of crosscorrelation functions and Asymptotic estimates for 256 point Linear FM Up and Down-sweeps, with and without 10% Raised cosine window.



III. Geometric Interpretation

In this section we show that a simple geometric interpretation may be applied to the asymptotic results for the cross-correlation function. The geometric interpretation can be very useful in calculating the crosscorrelation function. In particular we show that asymptotic result is obtained directly from the area of a parallelogram in the time-frequency plane.

We begin by defining a parallelogram in terms of two sets of parallel lines in the t-ω plane. Figure 7. shows the 4 lines.

The area of the parallelogram is

$$\text{Area} = xy \sin \Theta \tag{15}$$

We must calculate the three quantities separately. To obtain $\sin \Theta$ we observe that

$$\text{Tan}\,\alpha = A, \quad \text{Tan}\,\beta = B.$$

Note that ω_d is defined with a negative slope. Thus we have

$$\text{Tan}(90°-\alpha) = \frac{1}{A}, \quad \text{Tan}(90°-\beta) = \frac{1}{B}$$

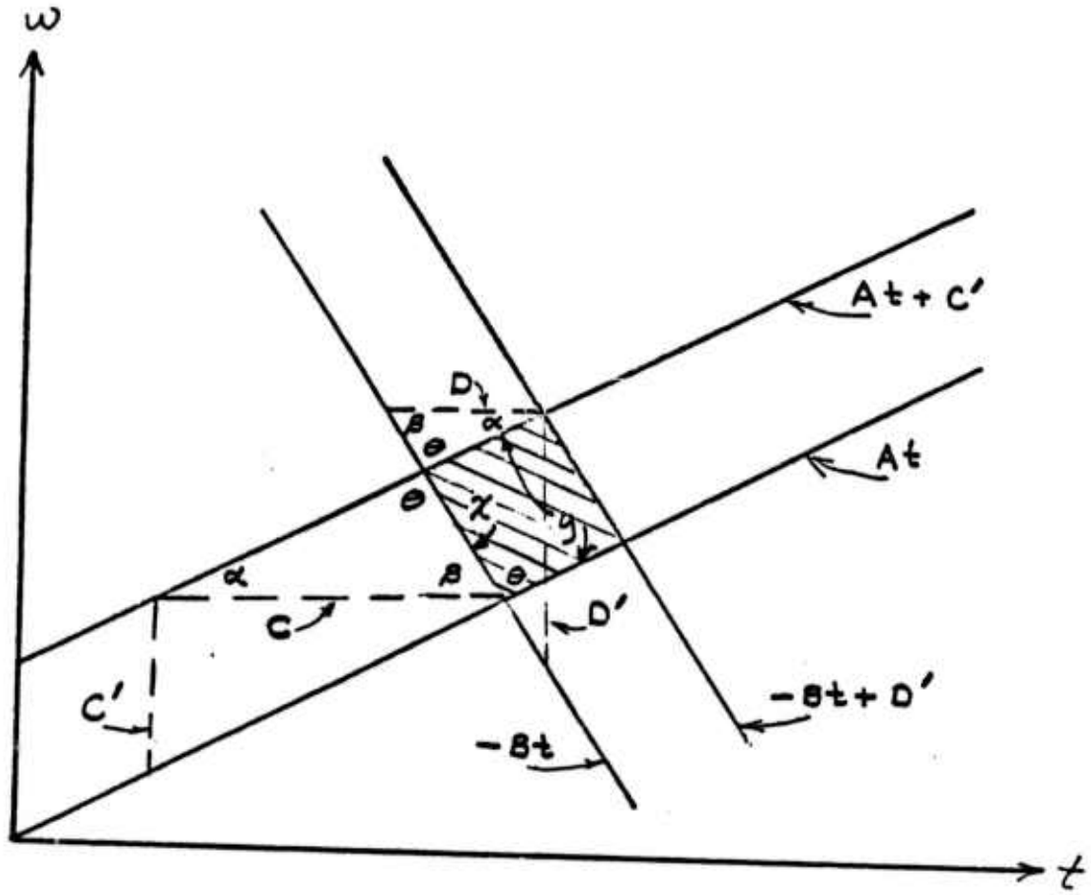
Taking inverse tangents and adding we have

$$\text{Tan}^{-1}\left(\frac{1}{A}\right) + \text{Tan}^{-1}\left(\frac{1}{B}\right) = 180 - \alpha - \beta = \Theta$$

Thus

$$\text{Tan}\,\Theta = \text{Tan}\left[\text{Tan}^{-1}\left(\frac{1}{A}\right) + \text{Tan}^{-1}\left(\frac{1}{B}\right)\right] = \frac{\frac{1}{A} + \frac{1}{B}}{1-\left(\frac{1}{A}\right)\left(\frac{1}{B}\right)} = \frac{A}{AB} + \frac{B}{-1}$$

Figure 7. Geometry for calculating the Area of Parallelogram enclosed by two sets of parallel lines.



and from this we obtain

$$\sin\theta = \frac{A + B}{\sqrt{(A+B)^2+(AB-1)^2}}$$

$$\sin\theta = \frac{A + B}{\sqrt{(A^2+1)(B^2+1)}} \qquad (16)$$

To obtain x and y we recall that the diameter of the inscribed circle for a triangle is calculated as the ratio of any side to the sin of the opposite angle. Thus

$$\frac{C}{\sin\theta} = \frac{x}{\sin\alpha}, \quad \frac{D}{\sin\theta} = \frac{y}{\sin\beta}$$

or

$$x = \frac{C\sin\alpha}{\sin\theta}, \quad y = \frac{D\sin\beta}{\sin\theta} \qquad (17)$$

but

$$\sin\alpha = \frac{A}{\sqrt{A^2+1}}, \quad \sin\beta = \frac{B}{\sqrt{B^2+1}} \qquad (18)$$

Thus combining (16), (17), and (18) and substituting into (15) we obtain

$$\text{Area} = \frac{CD(AB)}{A+B} \qquad (19)$$

Finally we observe that

$$\tan\alpha = A = \frac{C'}{C}, \quad \tan\beta = B = \frac{D'}{D}$$

and substituting we have

$$\text{Area} = \frac{C'D'}{A+B} \tag{20}$$

Here C' and D' are the respective vertical lengths between the two sets of parallel lines and A and B are the tangents of the lines (one positive and one negative).

Consider now the magnitude - squared of equation (13), we obtain

$$|R_{fg}(\tau)|^2 \sim \frac{2\pi}{\chi \frac{d^2h(\hat{t},\tau)}{dt^2}} \; k^2(\hat{t},\tau) \tag{21}$$

which is

$$|R_{fg}(\tau)|^2 \sim \frac{2\pi}{\frac{d\omega_f(\hat{t},\tau)}{dt} - \frac{d\omega_g(\hat{t},\tau)}{dt}} \; a^2(\hat{t},\tau) \; b^2(\hat{t},\tau) \tag{22}$$

If we now assume that the two instantaneous frequency functions are slowly varying, i.e. that we may replace ω_f and ω_g by the first two terms of their Taylor series expantion about the point $\hat{t}$, for fixed τ, then we may observe that if

$$C' = a^2(\hat{t},\tau)$$
$$D' = b^2(\hat{t},\tau)$$

which are also assumed to be slowly varying, and

(1.19)

$$A = \frac{1}{2\pi} \frac{d\omega_f(\hat{t},\tau)}{dt}$$

$$B = -\frac{1}{2\pi} \frac{d\omega_g(\hat{t},\tau)}{dt}$$

Then we have that

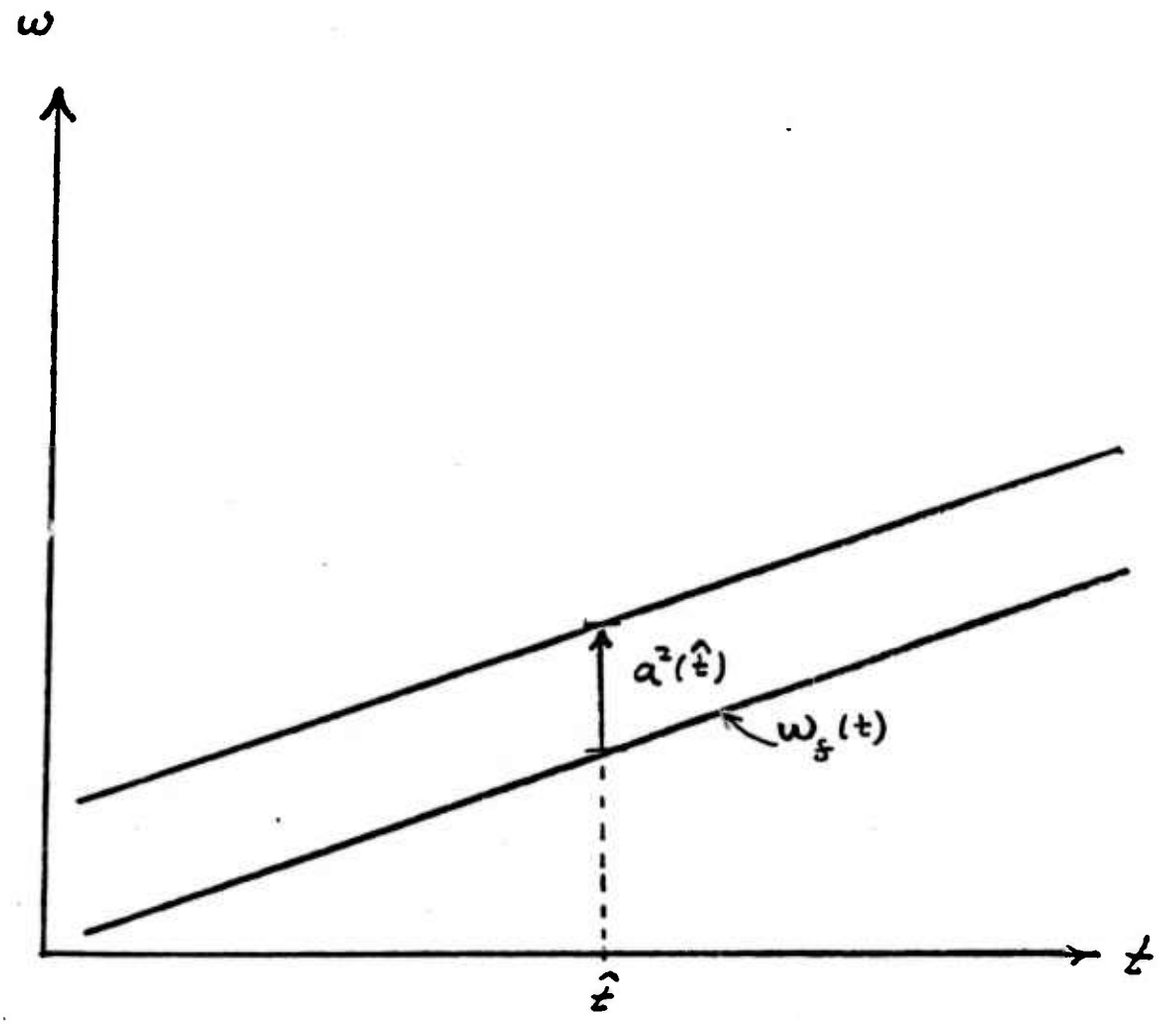
$$|R_{fg}(\tau)| \sim [\text{Area}]^{1/2} \qquad (23)$$

and the crosscorrelation function is asymptotically the square-root of the intersecting parallelogram described above. The results are clearly true for the two linearly frequency modulated signals provided in the example, since all the assumption concerning slow variations are true. Although we, as of yet, have not been able to prove the result when we remove the slow variation assumptions we feel the result is still approximately valid. Thus, one could construct a template for each signal such as seen in Figure 8.

Then overlaying the two signal templates, and calculating the square-root of the overlapping area, we obtain an estimate of the cross correlation function for that value of time difference. Moving the template horizontally would yield a new value of crosscorrelation.

Although it may seem confusing as to why we plot the amplitude-squared along the frequency axis, i.e. $C' = a^2(\hat{t},\tau)$, observe that for the rectangular amplitude case, as in the example, $a^2(\hat{t},\tau) = 1/T$ which has units of frequency, so that the results are consistent.

Figure 8. Template for a typical Linear FM Upsweep signal with amplitude a(t).



IV. Multiple Intersections

The case of multiple intersections and signals is considerably more complicated. The reason for this becomes clear when we consider a simplified case. Suppose that for a particular delay, the instantaneous frequency curves have two intersections. Then we may calculate the crosscorrelation contribution from each intersection, but since the actual correlation is the complex sum of the two since the phases must be considered. In principle we have the asymptotic expression for the phases in equation (13) and could thus calculate the complex sum of the two contributions. But if the two signals are only specified by their time-frequency plots and amplitudes we lose control of absolute phase. To show this, suppose that g(t) is the same as equation (6-b) and f(t) is

$$f(t) = a_1(t)e^{jA_1\alpha_i(t)} + a_2(t)e^{jA_2\alpha(t)} \tag{24}$$

or

$$f(t) = f_1(t) + f_2(t) \tag{25}$$

This could represent a signal with a fundamental and second harmonic component, where, for example, the second phase could be

$$A_2\alpha_2(t) = 2[A_1\alpha_1(t)]$$

then

$$R_{fg}(\tau) = R_{f_1g}(\tau) + R_{f_2g}(\tau). \tag{26}$$

Using stationary phase once again for each of the two integrals, the two stationary points $\hat{t}_1$ and $\hat{t}_2$, are the unique solutions to the equations

$$w_{f_1}(\hat{t}_1) - w_g(\hat{t}_1) = 0 \tag{27-a}$$

and

$$w_{f_2}(\hat{t}_2) - w_g(\hat{t}_2) = 0 \tag{27-b}$$

with

$$\chi_1 h_1(t,\tau) = A_1\alpha_1(t) + B\beta(t) \tag{28-a}$$

$$\chi_2 h_2(t,\tau) = A_2\alpha_2(t) + B\beta(t) \tag{28-b}$$

and

$$k_1(t,\tau) = a_1(t)b(t+\tau) \tag{29-a}$$

$$k_2(t,\tau) = a_2(t)b(t+\tau) \tag{29-b}$$

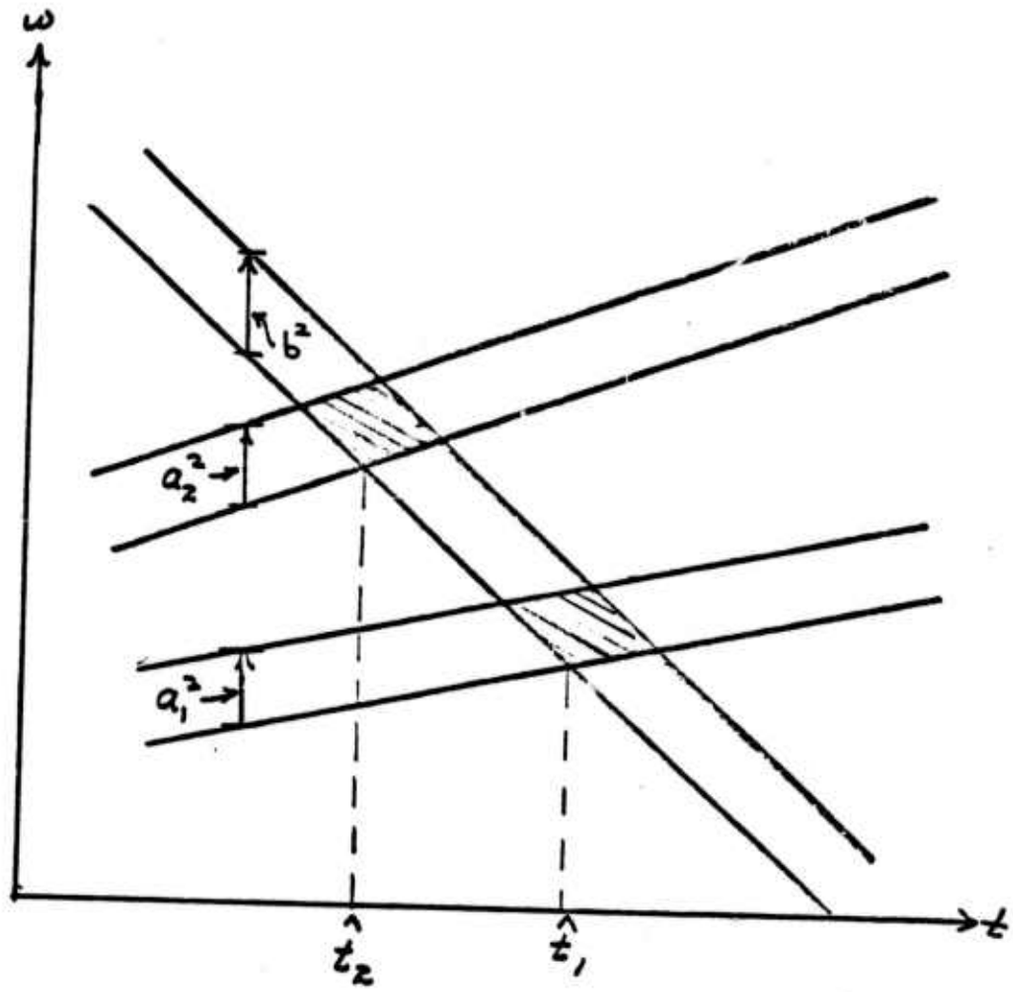
we have

$$R_{fg}(\tau) \sim \left[\frac{2\pi}{\chi_1 \dfrac{d^2h_1(f_1,\tau)}{dt}}\right]^{1/2} k_1(\hat{t},\tau)\exp[j\chi_1 h_1(\hat{t}_1,\tau) + j\pi/4] + \tag{30}$$

$$+ \left[\frac{2\pi}{\chi_2 \dfrac{d^2h_2(\hat{t}_2,\tau)}{dt^2}}\right]^{1/2} k_2(\hat{t}_2,\tau)\exp[j\chi_2 h_2(\hat{t}_2,\tau) + j\pi/4]$$

Figure 9. shows a typical time-frequency plot.

Figure 9. Overlapping templates for two intersection example.



The two shaded areas are the square of the amplitudes of the two contributions. Since we do not know, in general, what is the absolute phase relationship between the two components we can only say that crosscorrelation lies between the limits

$$|(\text{Area } 1)^{1/2} \pm (\text{Area } 2)^{1/2}| \qquad (31)$$

Clearly an upperbound on the crosscorrelation would be the sum of the square roots of the areas. Clearly, in more general cases, an upperbound to the crosscorrelation would be the sum of the square-roots of all intersecting areas. This would be accurate only if all phases were the same.

V. Crossabmiguity Functions

The results of the previous sections can be extended to wideband crossambiguity functions. In order to accomplish this consider

$$A_{fg}(\tau,s) = \sqrt{s} \int_{-\infty}^{\infty} f(t)g^*[s(t+\tau)]dt \qquad (32)$$

which is the wideband crossambiguity function for the two signals, $f(t)$ and $g(t)$. The parameter s is the Doppler stretch factor. If we assume that the functions are defined the same as in equations (6-a) and (6-b) then we can observe directly that

$$k(t,\tau) = \sqrt{s}\, A(\cdot)\, b[s(t+\tau)] \qquad (33)$$

and

$$h(t,\tau) = A\alpha(t) + B\beta(st). \qquad (34)$$

The stationary phase point occurs at value of t which is the solution of

$$w_f'(t) - s\, w_g'(\hat{st}) = 0 \qquad (35)$$

Thus equations (33) and (34) are substituted directly into equation (13) in order to obtain the asymptotic estimate of A_{fg}.

If, as in the example of linear FM,

$$w_f(t) = At, \; w_g(t) = A(T-t)$$

then the stationary point occurs at

$$\hat{t} = \frac{s(t-\tau)}{1 + s^2}$$

It should be observed, for this example, that the instantaneous frequency line for $\sqrt{s}\, g[st]$ is

$$w_{\tilde{g}}(t) = -A\, s^2 t + SAT$$

and the amplitude of g is modified by the stretch factor as well.

With these modifications in the functions h and k, the geometric procedure is directly applicable to the asymptotic estimate of crossambiguity functions. Clearly all comments and results for the multiple intersection case are equally valid.

VI. Conclusions

We have shown that a simple geometric procedure can be used to obtain the asymptotic estimate of crosscorrelation functions for signals whose instantaneous frequency curves cross at one point only. In other cases this procedure produces an upper-bound on the crosscorrelation function. The procedures described can be applied directly to ambiguity functions since it is a particular form of crosscorrelation function.

The particular utility of the geometric procedure lies in cases with which large numbers of FM signals may be checked for their correlation properties. Further as new electronic devices, which can represent signals in t-w space, are developed the area concept may proove useful as an identifier or matched filter.

Appendix

In this appendix we establish that if we have two analytic signals, k(t) and h(t), then the crosscorrelation of the real parts is equal to the real part of their crosscorrelation function. Further that the magnitude of the complex crosscorrelation function is the envelope of the real crosscorrelation. Finally, we give a computer example demonstrating this fact.

To begin, we assume that $f_1(t)$ and $f_2(t)$ are real, unit energy signals, and the transform

$$\hat{f}(t) = H[f](t) = \frac{1}{\pi} \int_{-\infty}^{\infty} \frac{f(\tau)}{t-\tau} \, d\tau \qquad \text{(A-1)}$$

is the Hilbert transform. Then we form the analytic signals

$$\begin{cases} k(t) = \frac{1}{\sqrt{2}} [f_1(t) + j \hat{f}_1(t)] & \text{(A-2-a)} \\ \\ h(t) = \frac{1}{\sqrt{2}} [f_2(t) + j \hat{f}_2(t)] & \text{(A-2-b)} \end{cases}$$

Defining the inner product at

$$(k,h) = \int_{-\infty}^{\infty} k(t) h^*(t) dt \qquad \text{(A-3)}$$

and using the properties of Hilbert transforms and their spectra is can be shown that if

$$\begin{cases} F_1(\omega) = A_1(\omega)e^{j\theta_1(\omega)} & \text{(A-4-a)} \\ F_2(\omega) = A_2(\omega)e^{j\theta_2(\omega)} & \text{(A-4-b)} \end{cases}$$

then

$$(k,h) = (f_1,f_2) + j[\frac{1}{\pi}\int_0^{\infty} A_1(\omega)A_2(\omega)\sin[\alpha(\omega)]d\omega] \qquad \text{(A-5)}$$

where

$$\alpha(\omega) = \theta_1(\omega) - \theta_2(\omega)$$

Recognizing that if

$$\begin{cases} f_1(t) = f(t) & \text{(A-6-a)} \\ f_2(t) = g(t+\tau) & \text{(A-6-b)} \end{cases}$$

Then the inner product is a crosscorrelation function and

$$R_e\{R_{kh}(\tau)\} = R_{fg}(\tau) \qquad \text{(A-7)}$$

Further since R_{kh} has a one sided spectrum then its real and imaginary parts are themselves Hilbert transforms and hence $|R_{kh}(\tau)|$ is the envelope of $R_{fg}(\tau)$.

To demonstrate this figure 10. shows the magnitude of the complex crosscorrelation and the crosscorrelation of the real parts of the up and down linearly sweeped signals shown in figures 2. and 3.,

Figure 10. Complex crosscorrelation and real cross-correlation of 128 point Linear FM Example, with 10% raised cosine window.

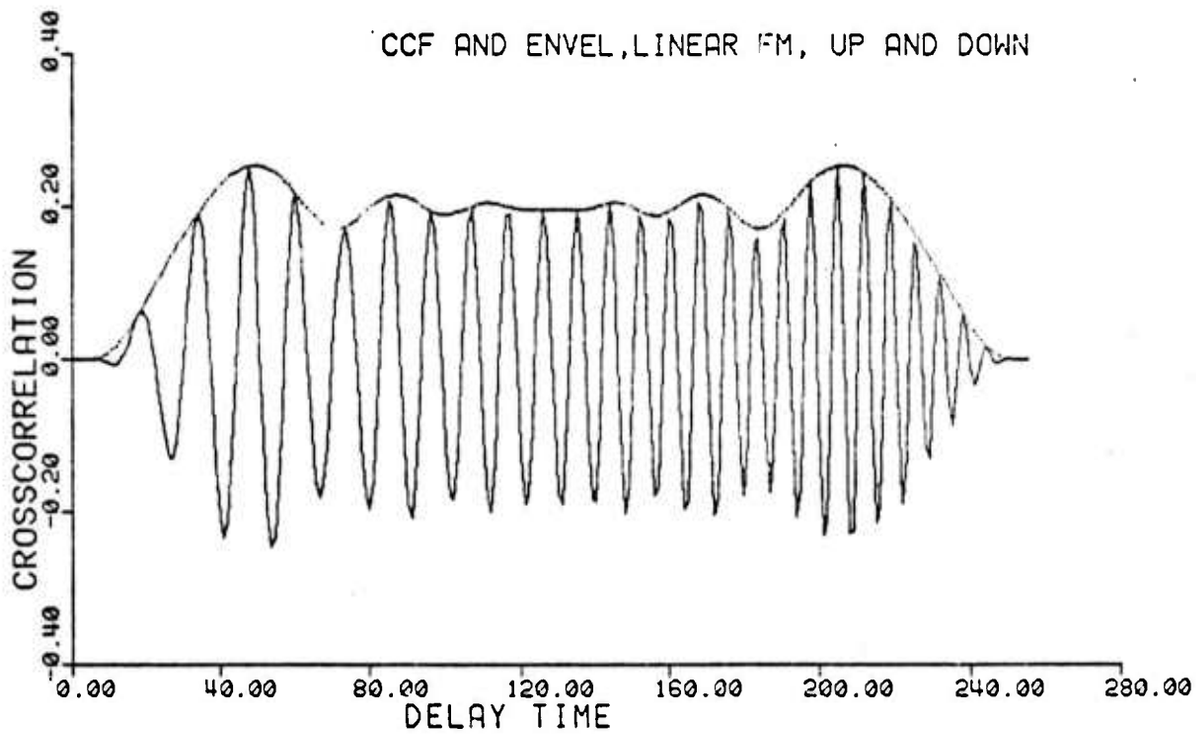
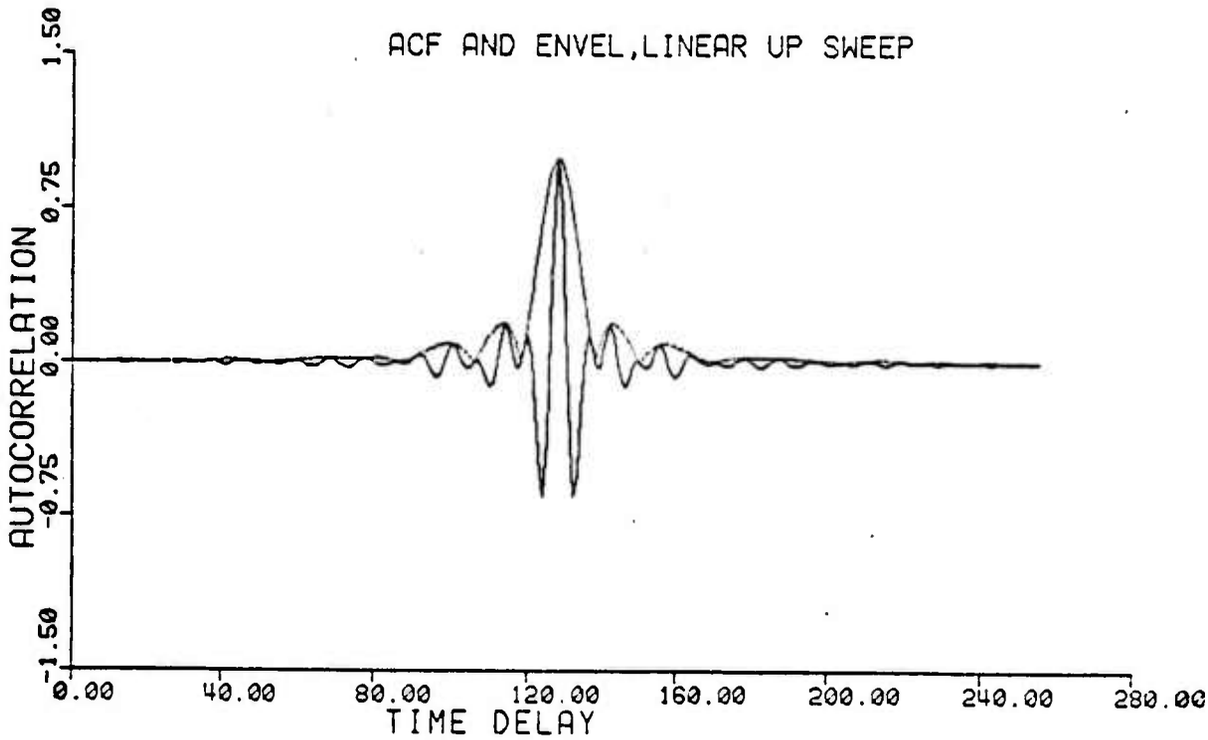


Figure 11. Complex autocorrelation and real autocorrelation of 128 Linear FM Upsweep, with 10% raised cosine window.



plotted on the same graph. Figure 11., shows the same results for the autocorrelation functions of the up sweep signal. The fact that real signal does not touch the envelope at several cycles is an artifact of the plotting routine drawing straight lines through two sample points, neither of which are at the peak value of the cycle.

Reference

1. Erdelyi, A., Asymptotic Expansions, Dover Publications, Inc., 1956.

Wideband Energy Density Functions

by

Edward L. Titlebaum
Associate Professor of Electrical Engineering
University of Rochester
Rochester, NY 14627

I. Wideband Energy Density Functions

In this section we outline our attempts at trying to find a definition of Energy density function for wideband signals, for which the Doppler effect is not a simple translation of the spectrum but must be considered a stretching of the signal. The procedure is to look for a transformation, T characterized by a kernel $k(t,p)$ so that if $\hat{f}(t)$ is some other transformation of a signal $f(t)$, and

$$\hat{F}(p) = \int_{-\infty}^{\infty} \hat{f}(t)k(t,p)dt \tag{1}$$

Then the function

$$A_f(t,P) = \hat{f}(t)\ \hat{F}^*(P)k(t,P) \tag{2}$$

would have some or all of the properties of the narrow band Energy density function[1,2]. Most of the transformations are based upon combinations of Fourier and Mellon Transforms. The reason for this is that the Fourier transform converts a time delay to an exponential multiplier in the transform domain and the Mellon transform converts time stretch into an exponential Multiplier in the transform domain[3,4]. In order to see this consider a signal $f(t)$ and its Mellon transform $\tilde{F}(x)$. Thus

$$\tilde{F}(x) = M[f(t)](x) = \int_{-\infty}^{\infty} f(e^t)\ e^{jxt}dt. \tag{3}$$

Inverting, we obtain that

$$f(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \tilde{F}(x)e^{-jx\ln(t)}dx \tag{4}$$

An alternate form for equation (3) is

$$\tilde{F}(x) = \int_0^\infty f(\tau) e^{-jx\ln\tau} d\tau/\tau \tag{5}$$

Now consider a stretch of f(t),

$$g(t) = f(st)$$

Then, from equation (4)

$$g(t) = \frac{1}{2\pi} \int_0^\infty \tilde{F}(x) e^{-jx\ln(st)} dt \tag{6}$$

from which we can easily see that

$$\tilde{G}(x) = \tilde{F}(x) e^{-jx\ln s} \tag{7}$$

which has the exponential factor. If we assume small stretching, ie.

$$S = 1 + \beta, \quad |\beta| << 1$$

then

$$\ln S = \ln(1+\beta) \tilde{\sim} \beta$$

and we have that

$$\tilde{G}(x) = \tilde{F}(x) e^{-jx\beta} \tag{8}$$

which is clearly the form we desire.

However, if we consider the effect of time-delay on the Mellon transform then we have that

$$f(t-\tau) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\tilde{F}(x)e^{-jx\ln(t-\tau)}dx \qquad (9)$$

and we observe that the kernel is <u>not</u> separable.

Suppose we try a Fourier-Mellon transform defined as[3]

$$G_f(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty}F(e^{\omega})e^{jx\omega}d\omega \qquad (10)$$

where F is the Fourier transform of f. Letting $p = e^{\omega}$ in equation 10, we obtain

$$G_f(x) = \frac{1}{2\pi}\int_{0}^{\infty}F(P)e^{-jx\ln P}dP/P \qquad (11)$$

as an alternate form for the F-M transform.

Now we consider a time delay of the signal[4]

$$h(t) = f(t-\tau)$$

Then

$$H(\omega) = F(\omega)e^{-j\omega\tau}$$

and

$$G_h(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty}F(e^{\omega})e^{-j\omega e}\ e^{jx\omega}d\omega \qquad (13)$$

or observing the Fourier form of this integral we observe that

$$F(e^{\omega})e^{-j\tau e^{\omega}} = \int_{-\infty}^{\infty} G_h(x)e^{-jx\omega}d\omega \tag{14}$$

where as

$$F(e^{\omega}) = \int_{-\infty}^{\infty} G_f(x)e^{-jx\omega}dx \tag{15}$$

We see that the only difference is an exponential factor.

Considering now a stretch of the signal so that again

$$g(t) = f(st)$$

and

$$G(\omega) = F(\omega/S)$$

which has F-M transform

$$G_g(x) = G_f(x)e^{jx\ln S} \tag{16}$$

and still retains the multilicative exponential form.

II. Possible Definitions of WBEDF

Based upon the form of equations (13) and (14) and the properties of the F-M transform, we define a wideband energy density function as

$$A_f(t,\tau) = f(t)F^*(e^P)e^{-jte^P}e^P \qquad (17)$$

We can show that this function has most of the integration properties of the narrow band EDF's. We assume that f(t) is a unit energy, analytic signal.

Properties of A_f

1. Integration with respect to P

$$\therefore \frac{1}{2\pi}\int_{-\infty}^{\infty} A_f(t,p)dp = \frac{f(t)}{2}\int_{-\infty}^{\infty} F^*(e^P)e^{-jte^P}e^P dp,$$

and letting $e^P = \omega$, or $p = \ln\omega$ and $dp = d\omega/\omega$. Thus we have

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} A_f(t,p)dp = \frac{f(t)}{2\pi}\int_{0}^{\infty} F^*(\omega)e^{-jt\omega}d\omega$$

Since we are dealing with analytic signals F has support (is non-zero) only for $\omega>0$ and we may extend the limits of integration from o to $(-\infty)$. Thus we have

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} A_f(t,p)dp = \frac{f(t)}{2\pi}\int_{-\infty}^{\infty} F^*(\omega)e^{-jt\omega}d\omega$$

(equation continued on next pg.)

$$= f(t)\ \frac{1}{2\pi}\int_{-\infty}^{\infty} F(\omega)e^{jt\omega}d\omega^{*}\ .$$

The bracketed term is f*(t), so that finally

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} A_f(t,p)dp = |f(t)|^2 \qquad (18)$$

2. Integration with respect to t

$$\int_{-\infty}^{\infty} A_f(t,P)dt = F^*(e^P)e^P \int_{-\infty}^{\infty} f(t)e^{-jte^P}dt$$

The integral is clearly F evaluated at e^{ω} so that

$$\int_{-\infty}^{\infty} A_f(t,p)dt = |F(e^P)|^2\ e^P \qquad (19)$$

We see that A_f has a positive t and p integrals.

3. Integration with respect to t and p

From either equation (18) or (19) we obtain

$$\frac{1}{2\pi}\int_{-\infty}^{\infty} A_f(t,p)dtdp = \int_{-\infty}^{\infty} |f(t)|^2 dt = 1 \qquad (20)$$

A second possible definition which retains the same integration properties as $A_f(t,p)$. Here we define $\tilde{A}_f(t,p)$

$$\tilde{A}_f(t,P) = G_F(t)F^*(e^P)e^{jtP}e^P \qquad (21)$$

Properties of $\tilde{A}_f$

1. $$\frac{1}{2\pi}\int \tilde{A}_f(t,P)dp = |G_F(t)|^2 \qquad (22)$$

2. $$\int \tilde{A}_f(t,P)dt = |F(e^P)|^2 e^P \qquad (23)$$

3. $$\frac{1}{2\pi}\iint \tilde{A}_f(t,P)dpdt = 1 \qquad (24)$$

The proofs of these properties are similar to the proofs for A_f, and will not be repeated.

III. Two dimensional Transform properties

One of the most important properties of the narrow band energy density function is its relationship to the narrowband ambiguity function. Thus, by extracting the two-dimensional Fourier transform of the NBEDF we obtain the signal ambiguity function. We will now establish that a similar transformation converts the WBEDF defined by $\tilde{A}_f(t,p)$ to the WB ambiguity function, except for the scale factor $\sqrt{s}$.

Consider the transform

$$T[\tilde{A}_f(t,P)](\tau,\lambda) = \frac{1}{2\pi}\iint_{-\infty}^{\infty}\tilde{A}_f(t,P)e^{j(t\lambda+\tau e^P)}dtdp \qquad (25)$$

If we recall that $\omega = e^P$ then this is a two-dimensional Fourier transform. Substituting for $\tilde{A}_f$ yields

$$T[\tilde{A}_f] = \frac{1}{2\pi}\int_{-\infty}^{\infty}F^*(e^P)e^P e^{je^P\tau}\int_{-\infty}^{\infty}[G_F(t)e^{jt(\lambda+P)}dt]dp \qquad (26)$$

The bracketed term in equation (26) is $F[e^{\lambda+P}]$ so that we have

$$T[\tilde{A}_f] = \frac{1}{2\pi}\int_{-\infty}^{\infty}F[e^{\lambda+P}]F^*(e^P)e^P e^{je^P\tau}dp \qquad (27)$$

Letting $\omega = e^P$ and $s = e^\lambda$, we obtain

$$T[\tilde{A}_f] = \frac{1}{2\pi}\int_{-\infty}^{\infty}F(s\omega)F^*(\omega)e^{j\tau\omega}d\omega \qquad (28)$$

Now substituting for F, we have

(2.9)

$$T[\tilde{A}_f] = \frac{1}{2\pi} \iiint_{-\infty}^{\infty} f(t_1) e^{js\omega t_1} f^*(t_2) e^{-j\omega t_2} e^{j\tau\omega} \, d\omega dt_1 dt_2 \quad (29)$$

Integrating with respect to ω and then t_2 yields

$$= \iint_{-\infty}^{\infty} f(t_1) f^*(t_2) \delta(t_2 - st_1 - \tau) dt_1 dt_2$$

$$= \int_{-\infty}^{\infty} f(t_1) f^*(st_1 + \tau) dt_1 \quad (30)$$

$$= \frac{1}{\sqrt{s}} \chi_f(\tau, s)$$

Equation (30) is the wideband ambiguity function for f(t), except for the square-root of s factor. Thus, if the kernel in (25) had been

$$k(t, \lambda, \tau, P) = e^{\lambda/2} e^{j(t\lambda + \tau e^P)}$$

we would have an exact transformation.

Clearly since the transform is based upon a Fourier Transform it is invertible and we can obtain $\tilde{A}_f$ from χ_f.

IV. Conclusions

We have been able to establish that certain of the properties of the narrowband energy density function we also valid for the two definitions of wideband energy density functions, through the use of the Fourier-Mellon transform. The important observation here is that signals should be represented in log frequency vs. time plots, and stretches in frequency become translations. The physical meaning of A_f and/or $\tilde{A}_f$ are subjects of continuing study, but we have established at least a link from the narrowband to the wideband theory of energy density functions.

V. References

1. Rihaczek, A. W., "Signal Energy Distribution in Time and Frequency," Trans. IEEE, IT-14, 1968, pp 369-374.

2. Ackroyd, M.H., "Short-time Spectra and Time-frequency Energy Distributions," JASA, 50, 1971, pp 1229-1231.

3. Moses, H. E. and Quesada, A. F., "The Power Spectrum of the Mellon Transform with Applications to Scaling of Physical Quantities," J. Math. Phys., Vol. 15, 1974, pp 748-752.

4. Altes, R. A., "The Fourier-Mellon Transforms and Mammalian Hearing," Unpublished manuscript.

AN EMPIRICAL INVESTIGATION OF THE APPLICATION OF MAXIMUM ENTROPY FILTERING FOR RESOLVING CLOSELY SEPARATED ECHOES

by

S. NARAYANA MURTHY

GRADUATE ASSISTANT

Dept. of Electrical Engineering

University of Rochester

Rochester, N.Y. 14627

Abstract

The possibility of using high resolution maximum entropy periodicity estimation technique to resolve closely separated echoes is studied in this report. By Fourier transforming the matched filter output, a short length periodic series in the frequency domain is obtained. The maximum entropy filtering method is applied to this series. A study of the relevant literature shows that tractable analytic models for determining the performance of the maximum entropy method with short length sinusoidal inputs do not exist. A limited number of computer simulations indicate that the technique does not resolve closely spaced echoes reliably.

Notation

Unless indicated otherwise, the following abbreviations and notations will be used throughout this chapter.

ACF : Autocorrelation function.

AR : Autoregression

ARMA : Autoregressive Moving average

B : Nyquist frequency

BW : 3 dB bandwidth

$\underline{C}$: $(C_1, C_2, \ldots C_L)^T$ = AR coefficients

DFT : Discrete Fourier Transform

$E\{\ \}$: Expectation operator

f : frequency variable, $-1/2 \leq f < 1/2$

$\underline{I}$: identity matrix

j : $\sqrt{-1}$

L : order of AR model

LMS algorithm : "Least mean square" algorithm [51]

N : Number of data samples

MEM : Maximum entropy method

PEF : Prediction Error Filter

PSD : Power spectral density

r.v. : random variable

$\gamma(k)$: $E\{x(i)\ x^*\ (i+k)\}$ = True ACF value of $\{x\}$ at lag K.

$\underline{\gamma}$: $(\gamma(1), \gamma(2), \ldots \gamma(L))^T$

$\hat{S}_{xx}(f)$: Estimate of the input PSD at frequency f by method xx

S(f) : "True" PSD of the input at frequency f; = Fourier Transform of $\{\gamma(k)\}$.

$\{x(i)\}_{i=1}^{N}$: set of N data samples

$\underline{X}_R$: $(x_R, x_{R-1}, \ldots x_{R-L+1})^T$

$\{n(i)\}_{i=1}^{N}$: set of independent additive noise samples.

α : adaption constant in LMS algorithm

$\delta_{j,k}$: Kronecker delta function

= 1 if j = k, 0 otherwise

ω : radian frequency = $2\pi f$.

Double underlining indicates a matrix; and single underlining, a vector.

superscripts: x* denotes complex conjugate of x.

$\underline{\underline{A}}^T$ " transpose of $\underline{\underline{A}}$

caret denotes estaimte, e.g. $\hat{\gamma}(k)$ = estimate of $\gamma(k)$

Section I.

Introduction

Underwater acoustic communication channels are often degraded by multipath and doppler scaling effects. It is believed [53] that the multipath propagation can be adequately modeled by a small number of point reflectors. If a signal $u(t)$ is transmitted into the channel, the received signal $r(t)$ can be represented by

$$\gamma(t) = \sum_{i=1}^{I} a_i \; u\,(S_i(t-\tau_i)) + n(t) \qquad (1.1)$$

where $\{\tau_i\}$ are the delays associated with the multipath, $\{S_i\}$ are the doppler scales, $\{a_i\}$ are the amplitudes and $n(t)$ is additive noise, usually considered to be white. The doppler effect can arise, for example, due to relative motion between transmitter and receiver.

Knowledge of $\{a_i\}$ and $\{\tau_i\}$ yield a better understanding of the nature of the communication channel, and are also useful for eliminating intersymbol interference effects.

The conventional technique for estimating $\{a_i\}$ and $\{\tau_i\}$ is to use a matched filter [48]. The filter matched to $s(t)$ is a linear system with impulse response given by $h(t) = s(T-t)$ where T is a delay usually included to make the filter realizable.

If the received waveform is processed with the matched filter for $u(t)$, the output is given by

$$x(t) = \sum_{i=1}^{I} a_i \, X_u \, (S_i, \tau_i) + \text{Noise Term} \qquad (1.2)$$

where $X_u(s_i, \tau_i)$ is the wideband ambiguity function of $u(t)$ defined by

$$X_u(S_i, \tau_i) = \sqrt{S_i} \int u(t) \, u(S_i(t-\tau_i)) \, dt \qquad (1.3)$$

The effect of doppler scaling can be made insignificant by choosing $u(t)$ to be a doppler tolerant signal. Such a signal is of the form

$$u(t) = a(t)e^{jk\ln t} , \qquad o \leq t \leq T , \quad \text{where}$$

$a(t)$ and K determine the bandwidth of $u(t)$ and $a(t)$ determines the sidelobe levels of $X_u(S_i, \tau_i)$. The doppler tolerant signal has the property that

$$X_u(S_i, \tau_i) \simeq e^{jk\ln S_i} \, X_u(1, \tau_i)$$

Thus the effect of using this type of signal is to convert the nonlinear doppler scaling effect to a simple linear phase shift. Detailed discussion of the ambiguity function and doppler tolerant signals can be found in [6].

With matched filter processing $\{a_i\}$ and $\{\tau_i\}$ are estimated as the largest I peaks of $x(t)$.

The closest separation between any two delays $\{\tau_i\}$ that can be detected by the matched filter is inversely proportional to

the bandwidth of u(t) (with suitable definition of bandwidth and resolution). In underwater acoustic channels, high frequency signal components are highly attenuated. The low bandwidth of the acoustic channel may limit the resolution provided by the matched filter to unacceptably low levels.

The Fourier Transform of x(t) in eq. (1.2) can be written, if u(t) is a doppler tolerant signal, as

$$X(f) = \sum a_i \, e^{jk\ln S_i} \; |U(f)|^2 \; e^{-j2\pi f\tau_i}$$

$$= |U(f)|^2 \; \sum a_i \, e^{jk\ln S_i} \; e^{-j2\pi f\tau_i} \qquad (1.4)$$

Equation (1.4) can be considered as a "time series" in frequency domain. The problem of estimating $\{a_i \, e^{jk\ln S_i}\}$ and $\{\tau_i\}$ are frequency domain analogs of the usual harmonic analysis of time series. However, since the bandwidth of $|U(f)|^2$ is low, the length of available data is very limited.

A nonlinear spectral analysis technique known as maximum entropy method has been proposed to obtain high resolution spectral estimates from small lengths of data. This report investigates the possibility of applying this technique to the resolution of closely spaced multipath delays.

Other nonlinear techniques are available for the estimation (and possibly, removal) of multipath delays. The most important among these are maximum likelihood spectral estimation methods [16,37] and cepstral analysis [10,54]. It is reported [14] that the former is not superior to the maximum entropy method.

No comparative study between cepstral analysis and maximum entropy analysis exists.

Section II.

Definition and Basic Properties of Autoregressive (AR) Spectral Estimators.

Unless otherwise indicated, the observed input will be assumed to be sampled at periodic intervals of 1 time unit and that only the sampled values are available for processing. Aliasing errors can be eliminated by lowpass filtering the analog input before sampling and such errors will be assumed to be negligible.

The conventional method of spectrum analysis is to estimate the autocorrelation function (ACF) of the observed data, multiply the estimated ACF by a suitable taper function and compute the Fourier transform of the tapered ACF. [9,39,54] This technique does not make use of any known structural properties of the observations. Consequently, one is forced to treat the spectral value at each frequency as an independent variable. Due to the large number of unknown quantities to be estimated, one requires large lengths of data in order to obtain adequate statistical stability.

Quite often, the input process can be satisfactorily approximated as the output of a discrete linear system driven by uncorrelated noise [1,12,13,24,27,32,35,43]. A special class of linear systems consists of systems which have only poles and no zeroes. In this case, the observations can be modeled by the stochastic difference equation

$$x(k) = \sum_{i=1}^{L} c_i \, x(k-i) + n(k) \qquad (2.1)$$

where $\{C_i\} = \underline{C}$ are unknown.

Equation (2.1) has the appearance of a regression equation in which the independent variables are past values of the observations themselves. Therefore, (2.1) is often called an autoregressive (AR) model of order L. For obvious reasons, it is called an all pole model also. Given x(k-1), x(k-2), ..., the best prediction of x(k) (in a least squares sense; and in a maximum likelihood sense if {n(k)} are gaussian) is a linear combination of the past input values. Hence, the use of the term "Linear Predictive Model". The terms {n(k)} are called residuals, prediction errors, or "innovations" since they represent the "new information" in {x(k)} , i.e., the part that cannot be predicted from past values. The filter with impulse response $(1, -C_1, -C_2, \ldots -C_L)$, is known as a prediction error filter (PEF), since the effect of operating on the input data with this filter is to obtain the prediction errors, {n(k)}. The PEF is sometimes called "whitening filter."

By multiplying both sides of (2.1) successively by n(k), x(k), x(k-1) .... x(k-L) and taking expectations, the following equations are obtained:

$$\begin{aligned}
&\gamma(0) - \gamma(1)\, C_1 - \ldots - \gamma(L)\, C_L = \sigma^2 \\
&\gamma(1) - \gamma(0)\, C_1 - \ldots - \gamma(L-1)\, C_L = 0 \\
&\qquad\qquad \vdots \\
&\gamma(L) - \gamma(L-1)\, C_1 - \ldots - \gamma(0)\, C_L = 0
\end{aligned} \tag{2.2}$$

from which the L+1 unknowns $\underline{C}$ and σ^2 can be estimated.

An alternative method of computing $\underline{C}$ is by solving the linear equations

$$\begin{Bmatrix} \gamma(0) & \cdots & \gamma(L-1) \\ & & \\ \gamma(L-1) & \cdots & \gamma(0) \end{Bmatrix} \begin{Bmatrix} C_1 \\ C_2 \\ \vdots \\ C_L \end{Bmatrix} = \begin{Bmatrix} \gamma(1) \\ \gamma(2) \\ \vdots \\ \gamma(L) \end{Bmatrix} \quad (2.3)$$

and σ^2 can be calcualted from the first equation of (2.2). Equations (2.3) are known as the normal equations, or Yule-Walker equations.

The same equations can be obtained by minimizing the mean squared prediction error

$$E\left(x(k) - \sum_{i=1}^{L} C_i x(k-i)\right)^2 \quad (2.4)$$

By applying a z-transform [54] to both sides of the first equation in (2.2), it can be seen that

Power Spectral Density (PSD) of the input = $S_{AR}(f)$

$$= \frac{\sigma^2}{\left|1 - \sum_i C_i\, e^{-j2\pi fi}\right|^2} \quad (2.5)$$

This equation provides a method of calculating the input PSD from a knowledge of the PEF coefficients. Since the PSD of the PEF is the inverse of the input PSD, the PEF is known as an inverse filter.

An alternative inpterpretation for the AR spectral estimation technique has been presented by Burg [15]. Burg observes that the loss of entropy, in an information theoretic sense, by passing (band limited) white noise through a linear system with frequency response C(f) can be written as

$$\Delta\varepsilon = - \ln |C(f)|^2 \, df \qquad (2.5)$$

This expression is minimized subject to the constraint that the inverse Fourier Transform of $|C(f)|^2$ should be equal to the estimated data ACF values up to lag L. The result of this is equations (2.3). Burg labels this spectral estimation method "Maximum Entropy Method."

Makhoul [31] observes that the AR spectral estimate minimizes the integrated spectral ratio

$$\int_{-1/2}^{1/2} \frac{S(f)}{\hat{S}_{AR}(f)} \, df$$

This interpretation is useful for modeling a selected portion of the input spectrum by an AR process.

The main attraction of the AR spectral estimation method is, of course, that it reduces the number of unknown parameters to the minimum, and therefore better satistical stability of the estimates can be obtained. Possible sources of error of this method are: (i) the presence of zeroes in the linear system which

generates the observations, (ii) wrong choice of order of model, L, (iii) presence of additive noise in the observations. Due to the highly nonlinear nature of the technique, it is extremely difficult to analyze its performance rigorously. In the succeeding sections, we shall summarize the current state of investigation of this topic.

Several applications of AR modeling have been discussed in [3,4,11,12,23,25,30,33,36,38,44].

Section III. Computational Algorithms.

There are at least three different computational algorithms for the calculation of the AR coefficients, $\underline{C}$. In each case, it is assumed that the mean (dc) component of the data is zero or has been subtracted out.

III.1. By solving the Yule-Walker Equations.

In this method, the autocorrelation values of the data up to lag L are estimated and the Yule-Walker equations (2.3) are solved by directly inverting the ACF matrix. There are two different methods of estimating the ACF. In method 1(a), the estimates are

$$\hat{\gamma}(j) = \frac{1}{N} \sum_{k=1}^{n-j} x_k \, x_{k+j} \, , \qquad 0 \leq j \leq L < N \qquad (3.1)$$

and in method 1(b),

$$\hat{\gamma}(j) = \frac{1}{N-j} \sum_{k=1}^{n-j} x_k \, x_{k+j} \, , \qquad 0 \leq j \leq L < N \qquad (3.2)$$

Equation (3.1) effectively assumes that the data is extended with zeroes. Eq. (3.2) gives an unbiased estimate of $\gamma(j)$, the true autocorrelation for lag j, if $\{ x_i \} \sum_{i=1}^{N}$ are assumed to be samples of a stationary stochastic process. Asymptotically, as $N \to \infty$, the two estimates become identical for all finite lags. However, for small N, the two estimates may have significantly different properties. The ACF matrix estimated from (3.1) is guaranteed to be positive definite. It is quite possible that

the ACF matrix corresponding to (3.2) may have zero or negative eigenvalues, leading to computational instability and absurd estimates of PSD.

The Levinson algorithm [29] is an efficient method of solving the Yule-Walker equations. Briefly, this algorithm iteratively generates solutions of order k from solutions of order k-1. The solution for order 1 is trivial and can be written down by inspection. When k=L, the algorithm terminates. The derivation the complex Levinson algorithm is presented in Appendix - I. This algorithm requires on the order of L^2 multiplications and 2L storage locations for temporary variables.

It is interesting to speculate whether a different recursive scheme will lead to an algorithm with a smaller number of multiplications. For example, can solutions of order k be generated from solutions of order k/2? This "divide and conquer" concept is responsible for the computational effectiveness of the Fast Fourier Transform algorithm [17]. Our investigations along this line have proved futile. However, it should be pointed out that since L is usually relatively small, the computational effort required for computing the ACF far outweighs the effort to invert the ACF matrix.

III.2. Burg's Algorithm

Burg [15] has proposed an algorithm for computing the MEM filter coefficients which does not require explicit estimation of the ACF. This method calculates the prediction error filter coefficients (p.6) by an iterative scheme that resembles Levinson's algorithm. A derivation of the complex Burg algorithm is given

in Appendix - II. Note the different error criterion, eq. A-2.8. This ensures that the computed PEF coefficients will represent a stable filter.

This algorithm requires $\simeq NL^2$ multiplications and $\simeq 2N$ storage locations. Clearly, this algorithm is most effective when the number of data samples, N, is small. When N becomes large compared to L, the Levinson algorithm and Burg algorithm tend to become identical.

III.3. Gradient Based Methods.

It is possible to calculate the AR coefficients without explicitly calculating or inverting the ACF matrix. The idea is to use some form of stochastic approximation (gradient seeking) method to solve eq. (2.3) [49,51,52]. The algorithms are of the general form

$$\underline{C}_{k+1} = \underline{C}_k + \alpha(k)\, \underline{X}_k\, (x(k+1) - \underline{X}_k^T\, \underline{C}_k) \qquad (3.3)$$

where $\underline{C}_k$ denotes the estimate of $\underline{C}$ at sampling instant k, $\alpha(k)$ is a predetermined gain sequence and

$$\underline{X}_k = (x(k), x(k-1), \ldots x(k-L+1))^T$$

The choice of $\alpha(k) = \alpha$ = constant corresponds to the Widrow-Hoff LMS algorithm [51]. The most striking feature of this algorithm is its simplicity, since only 2LN multiplications and L storage locations are necessary. This algorithm can also track slow variations in $\underline{C}$. The price paid for these advantages is that

relatively long data sequences are necessary to obtain satisfactory convergence of the estimates of $\underline{C}$, and even then these estimates are corrupted by the so-called "misadjustment noise" [51].

Section IV.

Statistical Properties of Spectral Estimators.

Before any statistical estimation procedure is applied in practice, it is useful to have an understanding of its statistical properties, such as its variance, probability distributions, confidence intervals, etc. In the case of PSD estimators in particular, intelligent tradeoffs between resolution, variability of the estimates and computational difficulty cannot be made without a good understanding of the behavior of the variable possible alternatives.

IV.1. Windowed DFT.

The statistical properties of the conventional windowed Discrete Fourier Transform techniques are well known and are discussed in detail in [7], [27], [9], [39], and [54]. Briefly, this method estimates the ACF values for lags 0, ..., N-1 via equations (3.1) or (3.2) and estimates the PSD by the relation

$$\hat{S}_{DFT}(f) = \sum_{k=-(n-1)}^{N-1} \hat{\gamma}(k)\, W(k)\, e^{-j2\pi fk} \qquad (4.1)$$

$$\hat{\gamma}(-k) = \hat{\gamma}(k). \qquad 0 \leq f \leq 1/2$$

W(k) is a suitably chosen window (or taper) function such that W(-k) = W(k). It is often required that W(k) be a positive semidefinite function, i.e.

$$\sum_{k=-(N-1)}^{N-1} W(k)\, e^{-j2\pi fk} \geq 0 \quad \text{for} \quad 0 \leq f \leq 1/2 \qquad (4.2)$$

The purpose of this is to ensure that the resulting PSD estimate is nonnegative, as it should be. Some of the commonly used window functions are the triangular (Bartlett) window, the raised cosine, Hamming, Hanning, and Kaiser windows.

If N is a highly composite number, (i.e. it can be decomposed into a product of a large number of small prime numbers - for example, $N=2^k$), a fast algorithm known as the FFT can be used to compute $S_{DFT}(f)$ at discrete values of $f_m = m/N$. Fast algorithms for the computation of the ACF and PSD are given in [39] and [54].

Assume that $W(k) = 0$, $|k| > m$. Then if the true PSD of the random process under consideration is continuous, it can be shown that

$$\lim_{\substack{n\to\infty \\ n\to\infty \\ L/N\to 0}} \left| \hat{S}_{DFT}(f) - S(f) \right| \to 0 \text{ in probability} \qquad (4.3)$$

This property does not hold if the data contains strictly periodic components.

The effect of the window function is to locally average the power spectral density estimate that would be obtained if the window function were not used. This results in reducing the noise component of the PSD estimate at each frequency, while at the same time "smearing" sharp spectral lines into adjacent frequency cells. Therefore, the use of windows reduces the statistical variability of the PSD estimates at the expense of a reduction in resolution.

IV.2. AR spectral estimators.

If S(f) is assumed to be continuous and bounded away from zero and infinity, a result due to Berk [8] states that the AR spectral estimate is consistent.

$$\lim_{\substack{L\to\infty \\ N\to\infty \\ L/N\to 0}} \left| \hat{S}_{AR,L}(f) - S(f) \right| \to 0 \text{ in probability.} \qquad (4.4)$$

This result is intuitively obvious from the discussion in Sec. II. As $L\to\infty$, $N\to\infty$, $L/N\to 0$, the first L elements of the inverse Fourier Transform of $\hat{S}_{AR,L}(f)$ tend to the true ACF values, and if the tails of the ACF are negligibly small for sufficiently long lags, it is apparent that

$$\lim_{L\to\infty} \hat{S}_{AR,L}(f) \to S(f).$$

Berk further shows that

$$\sqrt{\frac{N}{L}}\,(\hat{S}_{AR,L}(f) - S(f))/S(f)$$

has a limiting normal distribution with mean zero and variance equal to 2 for $f \neq 0$ or $1/2$ and equal to 4 when $f = 0$ or $1/2$; and

$$\operatorname{cov}\left[\sqrt{\frac{N}{L}}\,(\hat{S}_{AR,L}(f_1) - S(f_1),\ \sqrt{\frac{N}{L}}\,(\hat{S}_{AR,L}(f_2) - S(f_2))\right] \qquad (4.5)$$

$$\to 0 \text{ in probability for } f_1 \neq f_2$$

These statements are valid only asymptotically. The behavior of the AR spectral estimator is seen tc be identical to that of the conventional windowed DFT spectral estimator, as $L, N \to \infty$ and $L/N \to 0$.

The Burg algorithm differs from the Levinson algorithm only in that the first and last L data samples are treated differently. Therefore, the Burg spectral estimates can be expected to have the same asymptotic properties as discussed above.

The AR PSD estimates using LMS algorithm cannot be consistent, no matter how large N is.

IV.3. Confidence Intervals for AR spectral Estimates.

Quite often, it is desirable to form interval estimates of parameters rather than point estimates. This is often done by setting up confidence intervals for the point estimtes. This topic is treated in detail in most texts on statistical inference, e.g. [41].

The derivation of useful confidence intervals of $\hat{S}_{AR}(f)$ has proved to be a very difficult, if not impossible, task. The problem was approached as follows:

Since the autoregressive coefficients $\underline{\hat{C}}$ have been obtained by conventional least squares regression analysis, in the large sample case, $\sqrt{N}\ \underline{\hat{C}}$ has the approximate multivariate normal distribution with mean $\underline{C}$ and covariance matrix $s^2\ \underline{A}_N^{-1}$ where

$$s^2 = \frac{1}{N - L} \sum_{i=1}^{N} (x_i - \underline{\hat{C}}^T x_{i-1})^2 \text{ and} \qquad (4.6)$$

(equation continued on next pg.)

$$\underline{A}_N = \frac{1}{N} \sum_{i=1}^{N} \underline{X}_{i-1} \underline{X}_{i-1}^T$$

Using this fact, the usual statistical inference on $\underline{C}$, such as hypothesis testing, confidence intervals, etc., can be carried out.

The quantity $(1 - \sum_{i=1}^{L} \hat{c}_i e^{-j2\pi fi})$ is also a Gaussian random variable whose mean and variance can be calculated in terms of those of $\hat{\underline{C}}$, viz., $\underline{C}$ and $s^2 \underline{A}_N^{-1}$. This follows from the fact that a linear combination of jointly gaussian random variables is again gaussian.

With suitable scaling, the quantity $\left|1 - \sum_{i=1}^{L} \hat{c}_i e^{-j2\pi fi}\right|^2$ can be shown to have a noncentral χ^2 distribution with 2 degrees of freedom. This quantity is, of course, the denominator of the expression for $\hat{S}_{AR}(f)$.

The quantity $s^2 = \hat{\gamma}(o) - \hat{C}^T \hat{\underline{\gamma}}$, appearing in the numerator of the expression for $\hat{S}_{AR}(f)$, is the estimate of the variance of the residuals. If the observations are gaussian, least squares regression theory shows that with proper scaling, s^2 is a χ^2 distributed random variable with N-L degrees of freedom. Moreover, the numerator and denominator of $\hat{S}_{AR}(f)$ can be shown to be independent, since, from least squares theory, s^2 and $\hat{\underline{C}}$ are independent. Therefore, at least in principle, the distribution of $\hat{S}_{AR}(f)$ can be calculated.

There are several difficulties in the practical application of the foregoing theory. First, there is no known closed form

expression for the distribution of the ratio of a χ^2 distributed r.v. to an independent noncentral χ^2 distributed r.v. More seriously, the unknown values of the true AR coefficients, namely $\underline{C}$, enter into the distribution of $\hat{S}_{AR}(f)$ in a nonlinear manner such that it cannot be factored or subtracted out easily. This restricts the utility of the preceding theory to simulation studies where $\underline{C}$ can be calculated exactly by analytic or computational means.

An approximate solution to the problem can be constructed from the expression for the asymptotic distribution of

$$\sqrt{\frac{N}{L}} \left(\frac{\hat{S}_{AR}(f) - S(f)}{S(f)} \right)$$

which is given in eq. (4.5). This expression is valid only if the number of observations is large, and it is not known how large the set of samples should be before this formula can be applied.

IV.4. Choice of the order of autoregression.

A good choice of L, the order of the AR model, is clearly important. If L is too small, the resulting AR fit will not represent the data very well in the sense of a small residual mean square error, while a large value of L wastes computational resources and may also lead to increased round off errors and other numerical problems.

There are several possible ways to choose an appropriate value of L. The physical mechanism which produces the observed data,

if known, may provide useful clues. Another possibility is to choose a large order, say L_1, AR fit to the observed data and calculate the distribution of the AR coefficients. The hypothesis that the true order of AR is L is equivalent to the hypothesis that

$$C_{L+1} = C_{L+2} = \cdots = C_{L_1} = 0.$$

This hypothesis can be tested by statistical inference methods [41]

Perhaps a simpler method is to fit AR models of different orders to the observed data and in each case, estimate the variances of the residuals. We have already indicated that the distribution of these variances can be related to a χ^2 distribution. Now the usual variance ratio tests can be applied to determine the lowest order beyond which increasing the order of AR does not result in a statistically signficant reduction in residual variance.

Akaik [4] has proposed an alternative which simplifies this procedure. He suggests a final prediction error (FPE) statistic

$$FPE(L) = \left(1 + \frac{L+1}{N}\right) s^2, \text{ with } s^2 \text{ as} \qquad (4.7)$$

in Sec. IV.3., eqn. (4.6)

The value of L which minimizes FPE(L) is taken to be the true order of autoregression.

This criterion is particularly simple to apply with gradient algorithms.

Simulation studies of the use of the FPE criterion in selecting the order of the AR model have been presented by Ulrych and Bishop [46]. FPE(L) for different values of L, using both the Yule-Walker solution and Burg algorithm, have been computed. The minimum attainable FPE for the Yule Walker solution is typically much lower than that for the Burg solution. This is not surprising, since the former was developed by minimizing the residual error energy. As L approaches N, the FPE for both methods increases considerably.

Section V. Generalizations.

V.1. Pole zero modeling.

A more general model for the observed data allows the white noise driven linear system generating the data to have poles and zeroes. The data are modeled by the difference equation

$$x(k+1) = \sum_{i=0}^{L} C_i\, x(k-i) + \sum_{i=0}^{M} b_i\, n(k-i) \qquad (5.1)$$

where the input white noise sequence $\{n(k)\}$ is unknown. $\{C_i\}$ and $\{b_i\}$ are the unknown system parameters to be estimated.

If $\{n(k)\}$ were known $\{C_i\}$ and $\{b_i\}$ can be easily estimated using classical regression methods. An alternative computation technique, using the LMS algorithm, has been proposed by Widrow et al [52] and called "Adaptive Noise Cancelling".

In statistical literature, the pole zero model (5.1) is also known as the Autoregressive Moving Average (ARMA) model or rational model.

Define the z-transform of a discrete sequence $\{x(i)\}$ by the relation

$$x(z) = \sum_{i} x(i)\, z^{-i} \qquad (5.2)$$

The region of convergence of this series will depend on the sequence $\{x(i)\}$.

With this notation,

$$X(z) = \frac{B(z)}{C(z)}\, N(z) \qquad (5.3)$$

where $$B(z) = \sum_{i=0}^{M} b_i \, z^{-i} \qquad (3.28)$$

$$C(z) = \sum_{i=0}^{L} C_i \, z^{-i}$$

The case where $B(z) = 1$ corresponds, of course, to an all pole model.

The problem of identifying $\{C_i\}$ and $\{b_i\}$ when $\{n(i)\}$ are unknown, is nonlinear and extremely difficult [7]. An approximate solution can be obtained as follows:

If all the zeroes of the numerator and denominator are within the unit-z circle, it is possible to express $1/B(z)$ as a Taylor series about $z^{-1} = 0$:

$$X(z) = \frac{1}{C(z) \cdot (1/B(z))} \cdot N(z)$$

$$= \frac{1}{\sum_{i=0}^{L} C_i \, z^{-i} \sum_{i=0}^{\infty} d_i \, z^{-i}} \cdot N(z) \qquad (5.4)$$

Both the polynomials in the denominator of (5.4) are analytic in the region $|z^{-1}|<1$ by virtue of the assumption that all the poles and zeroes of $X(z)$ are inside the unit circle. Therefore, (5.4) represents a stable infinite order AR process. It is reasonable to hope that the coefficients $\{d_i\}_{D+1}^{\infty}$ become negligible for sufficiently large values of D, so that (5.4) can be approximated by

$$X(z) = \frac{1}{\left(\sum_{i=0}^{L} C_i\, z^{-i}\right)\left(\sum_{i=0}^{D} d_i\, z^{-i}\right)} \qquad (5.5)$$

Spectral analysis of the all pole model (5.5) can be carried out quite easily. The estimation of the coefficients $\{C_i\}$ and $\{b_i\}$ in (5.1) poses a more difficult problem. One method, suggested by Graupe and Perl [20], is to estimate the coefficients of the equivalent AR model (5.5), use these estimates to calculate the residuals $\{n(k)\}$ and use the estimated residuals to compute $\{C_i\}$ and $\{b_i\}$ by regression methods. An attempt at estimating the error involved in this procedure has been made in [19] and [20], but the expressions for bounds on errors are unilluminating.

An alternative method, due to Durbin, is given in Anderson [7].

An application of all zero modeling, in data communications, is to the problem of eliminating intersymbol interference. If the information symbols $\{S(k)\}_{-\infty}^{\infty}$, usually assumed to be independent, identically distributed random variables with a finite, discrete support, is transmitted over a communications channel with impulse response $\{h_k\}_{o}^{L}$, the received symbols $\{x(k)\}$ can be represented as

$$x(k) = \sum_{i=0}^{L} h_i\, S_{k-i} + n(k) \qquad (5.6)$$

The problem is to reconstruct $\{S_k\}$ given $\{x(k)\}$. One

possibility is to rewrite the all zero model (5.6) as an equivalent all pole model which can be identified. By passing the received symbols through the all pole inverting filter, $\{S_k\}$ can be recovered.

It is essential to the success of this scheme that the z-transform of the channel impulse response should have its zeroes within the unit z circle. In order to keep the order of the inverting filter low, it is also desirable that these zeroes do not lie close to the unit circle.

The condition that all the zeroes of H(z) should be within the unit circle can be relaxed if one is willing to tolerate a non-causal all pole model for the input, i.e. x(k) is expressed in terms of its own past and future values plus additive noise

$$x(k) = \sum_{i=-M_1}^{M_2} d_i \, x(k-i) + n(k)$$

It is still essential that H(z) should not have any zeroes close to, or on, the unit z circle. Communication channels with a small number of point target reflectors are common examples that do not satisfy this condition and are, therefore, unsuitable for the application of inverting filter models.

Several nonlinear algorithms, such as decision feedback equalization and the Viterbi algorithm for maximum likelihood sequence estimation, for the elimination of intersymbol interference can be found in a survey article by Proakis [38].

V.2. Effects of additive noise and nonwhite driving noise.

Sometimes the output of the all pole system to be modeled is corrupted by additive noise before it can be observed. The appropraite mathematical model is

$$y(k) = \sum_{i=1}^{L} C_i\, y(k-i) + n(k)$$

$$x(k) = y(k) + n_a(k) \qquad (5.7)$$

where, as before, $\{C_i\}$ are the unknown parameters to be estimated, $\{n(k)\}$ is the unknown white noise driving sequence and $\{n_a(k)\}$ is the unknown additive white sequence. $\{x(k)\}$ are the observations and $\{y(k)\}$ are the unobservable outputs of the linear system to be modeled.

If we model the observations by an AR process

$$x(k) = \sum_{i=1}^{L_1} \hat{C}_i\, x(k-i) + n_1(k)$$

the resulting normal equations for the estimation of $\hat{C}_i$ are seen to be

$$(\underline{\hat{R}}_{\chi} + \underline{\hat{R}}_{Na})\, \underline{\hat{C}} = \underline{\hat{\gamma}}_{\chi} \qquad (5.8)$$

The esimates of $\underline{\hat{C}}$ obtained from this will not equal $\underline{C}$, i.e. bias in the estimation of $\underline{C}$ is unavoidable.

The effect of this on the AR spectral estimate is not clear. With windowed DFT spectral estimates, the effect of additive noise is simply to add a constant value to the noise-free PSD, without affecting the "features" of the latter. This is not necessarily true with AR PSD estimates. So caution has to be exercised in modeling noisy data by an AR process.

Next we consider the case where there is no additive noise but the driving noise sequence $\{n(k)\}$ is non-white. In this case, the AR coefficient estimates based on least squares will be biased, and so, presumably, will the spectral estimates based on these estimates. If the ACF matrix of $\{n(k)\}$ is known, a Gauss-Markov parameter estimation scheme can be used instead of a least squares scheme can be used to get unbiased estimates [41].

Another solution is possible if the driving noise ACF is known to be much narrower than the signal ACF, i.e. the ACF of $\{x(k)\}$. A simple example of this is when $\{n(k)\}$ is generated by an all zero (moving average) process. If we assume the structure given by eq. (2.1), with $E(n(k)\ n(k+k_1)) \approx 0$ for $K_1 \geq L_1 \geq 1$ for some small value of L_1, the equations for the unbiased least squares estimation of $\underline{C}$ are given by

$$\begin{aligned}
\hat{\gamma}(L_1) &= C_1 \hat{\gamma}(L_1-1) + C_2 \hat{\gamma}(L_1-2) + \ldots + C_L \hat{\gamma}(L_1-L) \\
\hat{\gamma}(L_1+1) &= C_1 \hat{\gamma}(L_1) + C_2 \hat{\gamma}(L_1-1) + \ldots + C_L \hat{\gamma}(L_1-L+1) \\
&\vdots \\
\hat{\gamma}(L_1+L-1) &= C_1 \hat{\gamma}(L_1-L-2) + \ldots + C_L \hat{\gamma}(L_1-1)
\end{aligned} \qquad (5.9)$$

These equations enable one to determine the poles of the linear all pole system generating {x(k)}. However, this knowledge is not sufficient to estimate the PSD of {x(k)}.

Section VI.a. AR spectral estimation with sinusoidal inputs.

Recently published research literature indicates an interest in applying AR spectral modeling to time series consisting of sums of sinusoids and white additive noise [14,21,28,36,40,44, 46,52]. The motivation for this approach is that a sinusoid can be generated by inputting white noise to an all pole system whose poles are on the unit z circle. Unfortunately, this system is unstable, and the preceding discussion of AR spectral analysis is not applicable to this case.

One may still model the observed time series formally by an AR process, such that the poles of the PEF all lie on the unit-z circle at $z = e^{j\omega_o}$, where ω_o = radian frequency of the input sinusoid. The coefficients $\underline{C}$ can be computed by solving the normal equations (2.3) and the PSD estimate, from (2.5).

A simplified analysis of AR PSD estimation with sinusoidal inputs has been attempted by Lacoss [28] and later, by Widrow et al. [52]. For simplicity, first consider the case where $\{x(k)\}$ consists of a single complex sinusoid of frequency f_o, and let $\omega_o = 2\pi f_o$ be the normalized radian frequence. Define

$$\underline{u} = (1,\ e^{j\omega_o},\ e^{j2\omega_o},\ \ldots\ e^{j(L-1)\omega_o})^T \qquad (6.1)$$

then $\hat{\gamma}(k) = \sigma^2\delta_{k,k} + e^{jk\omega_o}$ where $\delta_{j,j}$ is the Kronecker delta function. The solution of the normal equations is

$$\underline{C} = e^{j\omega_o}\ \frac{\underline{u}}{\sigma^2 + L}\ , \text{ and}$$

(equation continued on next pg.)

$$S_{AR}(\omega) = \frac{\sigma^2 + L - 1}{\sigma^2 + L} \cdot \frac{1}{\left| 1 - \frac{1}{\sigma^2+L} \sum_{n=0}^{L-1} e^{-jn(\omega_o-\omega)} \right|^2} \qquad (6.2a)$$

From this expression, together with the assumptions that $L>>1$ and $\frac{L}{\sigma^2}>>1$, it can be shown that the height of the spectral peak at $\omega = \omega_o$ is nearly equal to

$$S_{AR}(\omega_o) = \frac{L^2}{\sigma^4} \qquad (6.2b)$$

and the 3-dB bandwidth of the spectral peak, defined as $\frac{d^2 S_{AR}(\omega)}{d\omega^2}\Big|_{\omega=\omega_o}$ can be shown to be

$$BW_{AR} = \sigma^2/\pi L^2 \qquad (6.2c)$$

The corresponding equations for the conventional Bartlett windowed DFT spectral estimates are

$$S_{DFT}(\omega) = \frac{1}{L}\left(1+\frac{1}{\sigma^2 L}\left|\sum_{n=0}^{L-1} e^{j(\omega_o-\omega)n}\right|^2\right) \qquad (6.3a)$$

$$S_B(\omega_o) = \frac{1}{\sigma^2} \qquad (6.3b)$$

$$BW_B = \sqrt{6}/\pi L \qquad (6.3c)$$

Comparison of equations (6.2) and (6.3) seems to indicate that the bandwidth of the spectral peak estimated via AR modeling is nearly σ^2/L times the bandwidth of the peak from DFT analysis. For example, for an input signal to noise ratio of 10dB ($\sigma^2=1/10$) and AR estimator order of 10 (L-10), the AR spectral peak is narrower than the DFT spectral peak by a factor of 100.

Several potential sources of trouble are ignored in this analysis. It has been assumed that the ACF matrix of the additive noise can be replaced by $\sigma^2 \underline{I}$. In fact, this idealization can be achieved in practice only if N is large. A more careful analysis of the performance of the AR spectral estimator should replace $\sigma^2 \underline{I}$ by $\sigma^2 \underline{I} + \underline{N}_a$, where $\underline{N}_a$ is a random matrix which accounts for the effects of imperfect estimates of the noise ACF matrix. The effects of $\underline{N}_a$ on the inverse of the data ACF matrix can be considerable. The condition ratio - i.e. the ratio of the largest eigenvalue to the smallest - of the data ACF matrix for noisy sinusoidal inputs is on the order of L/σ^2. When this ratio is large - the condition under which the AR spectral estimator is claimed to have high resolution - the effect of small random perturbation on the inverse of the ACF matrix can be large. Therefore, the high resolution of the AR frequency estimators is offset by large statistical variability.

The expression for the resolution of the AR spectral lines is misleading for a second reason. The use of

$$\left| \frac{d^2 S_{AR}(\omega)}{d\omega^2} \right|_{\omega=\omega_o}$$

as a measure of the resolution of the AR

spectral estimator is questionable. It is well known that the ACF of the sum of uncorrelated random variables is equal to the sum of the respective ACFs. For PSD estimators which perform linear operations on the estimated ACF, such as the windowed DFT methods, the use of the resolution measure given above is acceptable. However, the inverse of two ACF matrixes is not necessarily equal to the sum of the individual inverses. Therefore, the AR PSD estimates are not additive, and the bandwidth of one spectral line may be quite strongly influenced by the presence of another spectral line nearby.

Computer simulation appears to be the only proper method of evaluating the performance of AR spectral estimators with noisy sinusoidal inputs. Results presented in [40], [44], and [46] suggest that AR spectral line estimation can indeed give high resolution spectral estimates.

Section VII. Computer Simulation of delay estimation with matched filters and AR modeling.

In section I, we observed that the problem of estimating time delays is equivalent to the estimation of periodicites in frequency domain. As shown in previous sections, no adequate theoretical models exist for studying the behavior of AR frequency estimators with short lengths of data. Therefore, we had to resort to empirical computer simulation to study the problem.

A large number of variables can have possible effects on AR delay estimates. Initially, we judged that the following were the factors most likely to influence the estimates.

1. Signal to Noise ratio (SNR): This was defined as the signal energy divided by the noise variance. This definition is the "output SNR" used in evaluating matched filter performance. The noise was appropriately bandpass filtered so that its band occupancy was the same as that of the data. SNRs of 30, 20, and 10 dB were considered, representing low, moderate and high levels of additive noise.

2. The shape of the matched filter spectrum envelope: Gaussian, triangular and rectangular spectral envelopes were used. These correspond to waveforms with low, moderate and high side lobe levels in the time domain.

3. Number of data samples in the frequency domain. This quantity was defined as the number of frequency samples between 3-σ points in the case of gaussian envelopes, and the zeroes of the envelope in the triangular case. The variable was made to take on values of 32, 64, and 128 by appropriate choice of time domain pulse.

4. Location of delay: Since the AR technique is highly non-linear, it is conceivable that the delay estimates could depend on the actual location of the echoes. Three values of delay, namely 25, 64, and 125 were considered.

5. AR order: Values of 4, 8 and 16 were experimented with. Obviously, the order of the prediction error filter is one greater than the AR order.

The algorithm used for generating additive noise was of the linear recurrence type described in Abramowitz and Stegun [55]. A sequence of uniformly distributed random variables was generated by the relation $u_{k+1} = (a\ u_k + b) \bmod T$ where $a = 129$; $b = 1$; $T = 2^{35}$ and $u_o = 10987654321$.

After scaling by T, pairs of u(o,1) random variables were converted to gaussian r.v.s by the relation

$$n_1 = \sqrt{-2\ln u_1}\ \cos 2\pi\ u_2$$

$$n_2 = \sqrt{-2\ln u_1}\ \sin 2\pi\ u_2$$

Double precision (64 bit) arithmetic was used to generate these r.v.s.

The simulations were run on an Interdata 7/32 32-bit machine. The Burg algorithm and Levinson algorithms were computed with 64-bit precision, while the FFT calculations were made with 32-bit precision.

The results are presented in Figures 1(B) through 34(B) and 1(L) through 8(L). The subscripts B and L denote processing

with Burg and Levinson algorithms respectively.

It is hardly feasible to study all possible combinations of all the different values of the variables listed above. In order to make the investigation manageable, some of the variables were omitted from further consideration when it appeared that they did not influence the estimates much or when they degraded the performance significantly.

Figs. 3(B) and 6(B) show that at low SNR (=10 dB) low order (=4) AR estimates are too flat and the high order (=16) estimates show spurious peaks. Therefore, we did not consider the low SNR situation further.

Figs. 1(B) and 4(B), 2(B) and 5(B) and 9(B) and 11(B) show that for the same SNR, there is little difference between the estimates with gaussian and triangular envelopes. So it was decided to work further only with gaussian envelopes.

It is observed from fig. 1(B) through 8(B) that the AR delay estimates are indeed sharper than the matched filter estimates. Increasing the number of data samples to 128 (Fig.9(B), 10(B), 11(B)) do not affect the AR estimates to any great extent, while the matched filter estiamtes become sharper. When the number of samples is decreased to 32, (Fig. 12(B) and 13(B)) AR estimates of order 8 and 16 show noticeable bias and a tendency to split the single delay peak. This is in agreement with the observation made in Sec. IV that the variability of the AR coefficients increases sharply as the order of the AR model becomes comparable to the number of data samples. So long as the order of the model is much smaller than the number of data

samples, the AR estimates do not show much variation when the number of samples is changed. We should add the disclaimer that the few observations we have made are not sufficient to prove this observation.

Figs. 14(B) through 21(B) do not indicate that the AR delay estimate is significantly affected by the location of the delay, at least for small AR orders.

The conclusions drawn from Figs. 1(B) through 24(B) can be summarized as follows. The AR delay estimator does not perform well at low SNR or when the order of the AR model is greater than approximately one-fourth the number of available data samples. The location of the delay and the waveshape do not appear to have significant effects on the AR delay estimates.

Figs. 24(B) through 33(B) are the results of simulations of estimation of two delays. In figs. 24(B) and 28(B), both the matched filter and AR estimator show the presence of two distinct echoes. In Figs. 26(B), 27(B), 29(B) and 31(B), the matched filter shows two modes, indicating two echoes, but the AR estimators show only one peak. Finally, in Figs. 25(B), 28(B), 30(B), and 32(B) neither the matched filter nor the AR estimator is able to distinguish between the two peaks. The two modes of the 16th order AR output in Fig. 30(B) could well be spurious, as can be verified by comparing Fig. 30(B) and 22(B). After this series of failures, we did not feel it worthwhile to continue the simulations.

Fig. 1(L) through 8(L) do not indicate that any improvement is obtained by substituting the Levinson algorithm for the Burg algorithm.

FIG. 1(B)
DELAY ESTIMATION WITH BURG ALGORITHM & MATCHED FILTER; DATA DESCRIPTION : SNR = 30 DB; NUMBER OF DATA SAMPLES = 64; ACTUAL DELAY = 64; GAUSSIAN ENVELOPE; AR ORDERS 4, 8 & 16.

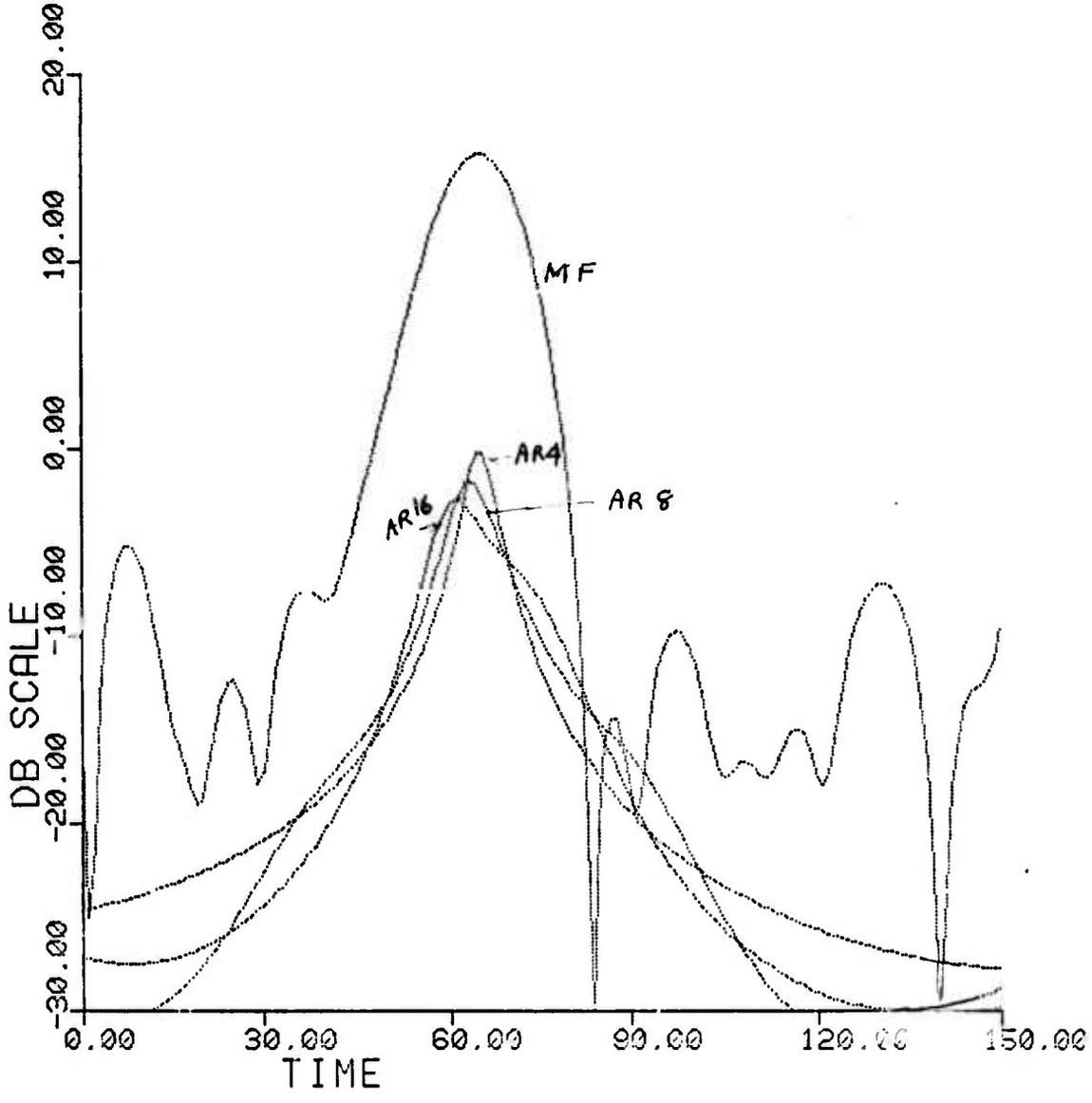
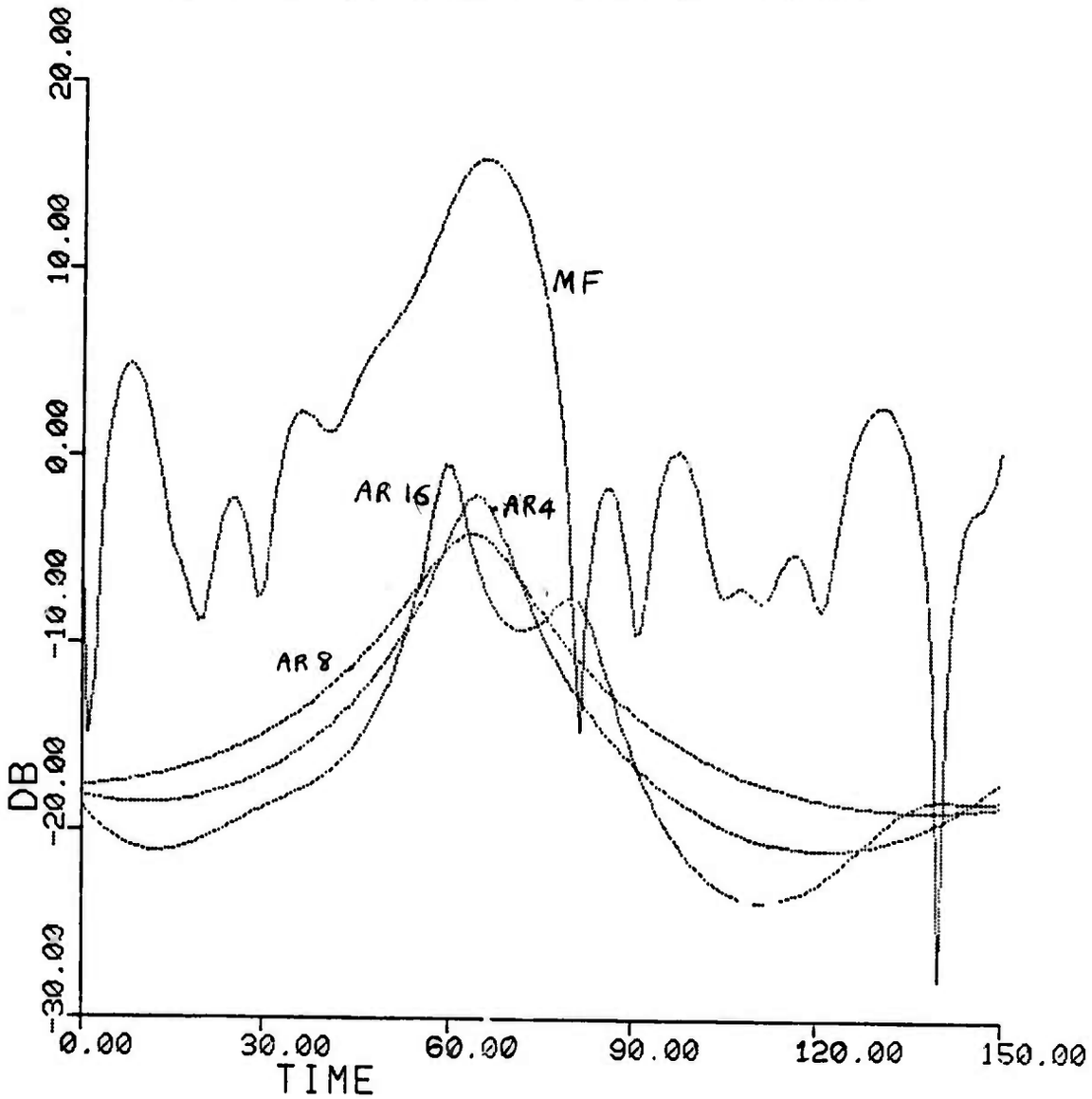


FIG. 2(B)
SAME AS FIG. 1(B), EXCEPT SNR = 20 DB.



FG. 3(B)
SAME AS FIG. 1(B), EXCEPT SNR = 10 DB.

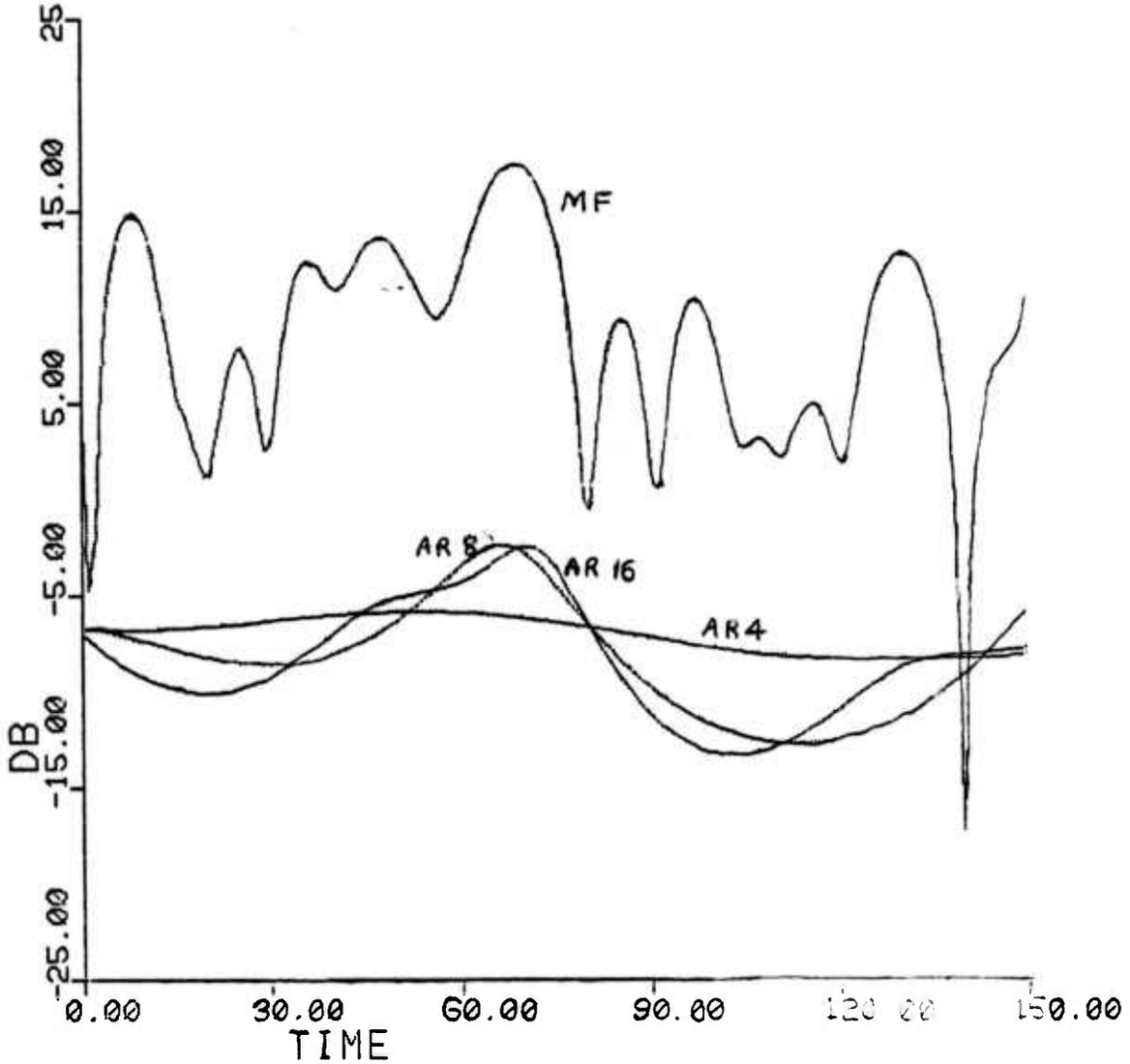


FIG. 4(B)
BURG ALGORITHM VS. MATCHED FILTER FOR DELAY ETIMATION : DATA DESCRIPTION ; SNR = 30 DB; TRIANGULAR ENVELOPE; NUMBER OF DATA SAMPLES = 64; ACTUAL DELAY = 64; AR ORDERS 4, 8 & 16.

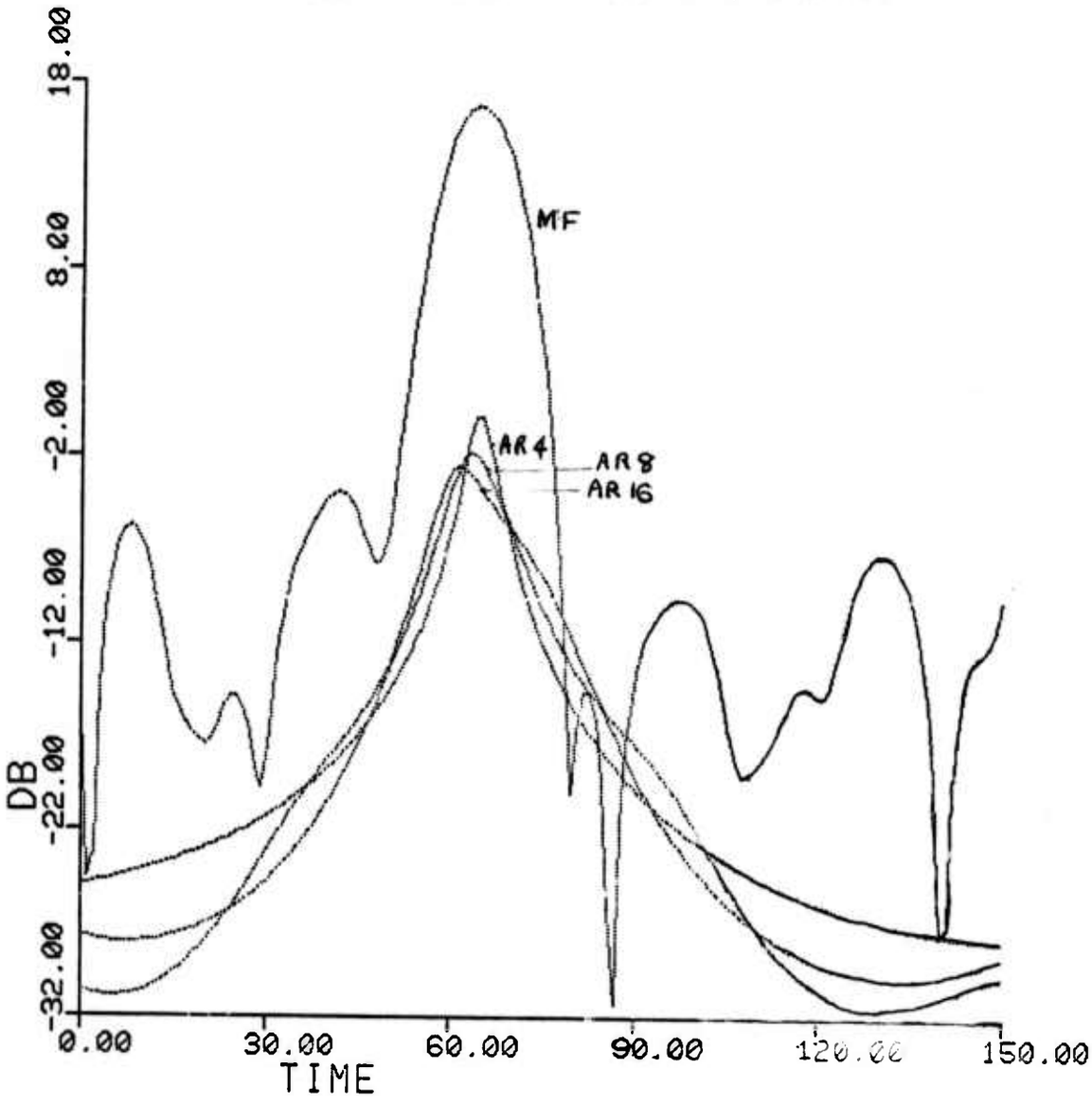


FIG. 5(B)
SAME AS FIG. 4(B), EXCEPT SNR = 20 DB.

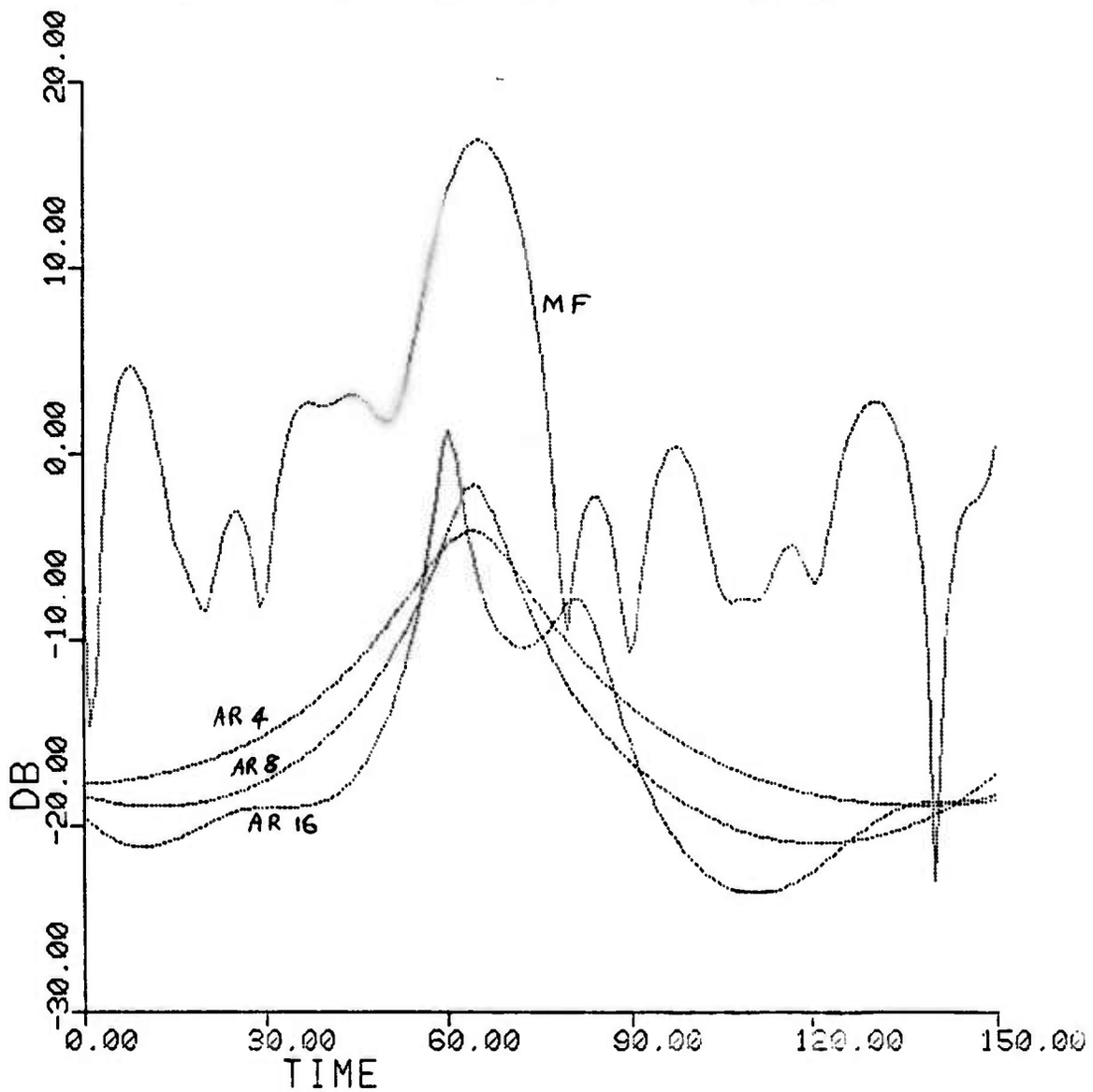


FIG. 6(B)
SAME AS FIG. 4(B) EXCEPT SNR = 10 DB.

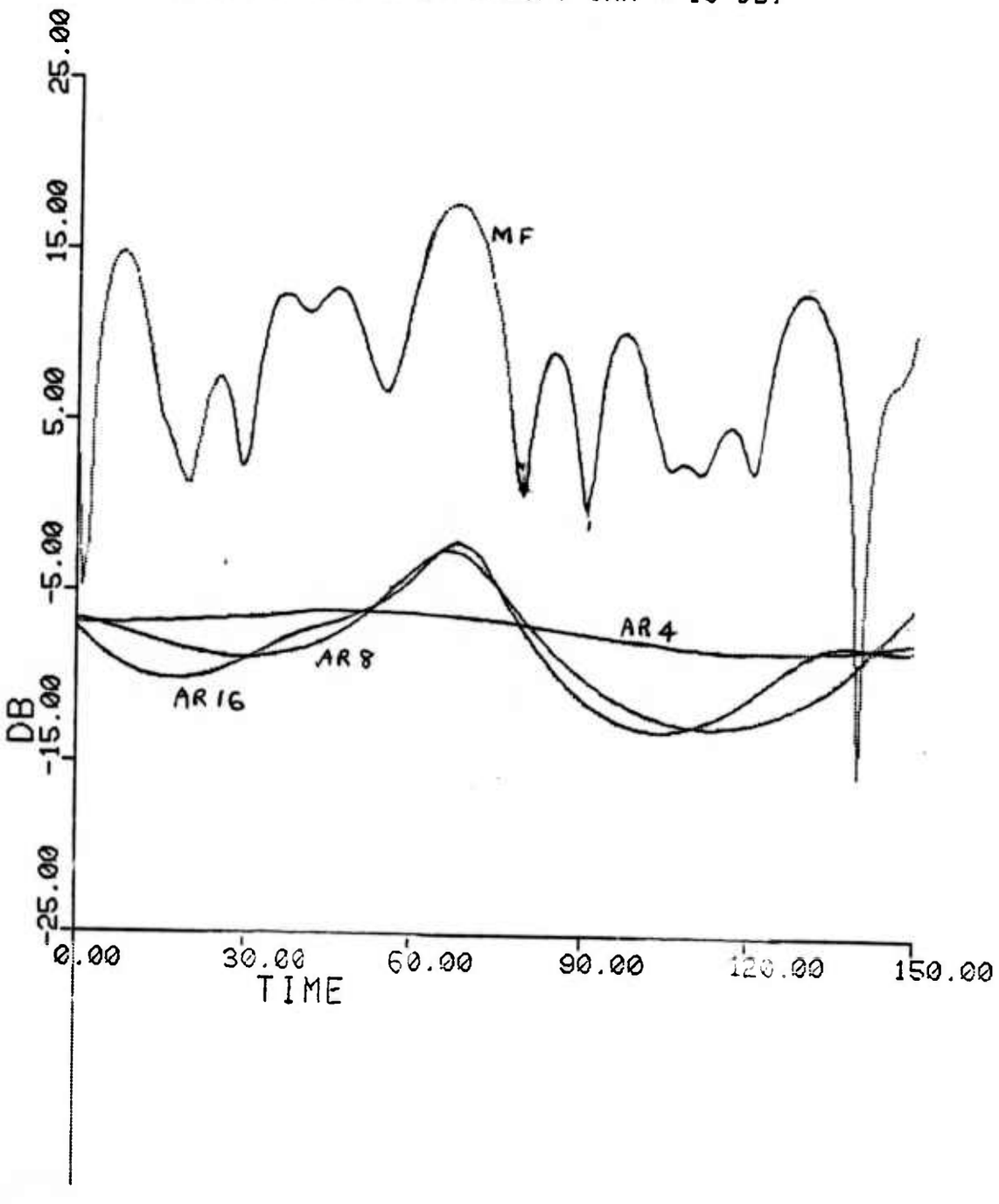


FIG. 7(B)
COMPARISON OF BURG ALGORITHM & MATCHED FILTER FOR DELAY ESTIMATION : SNR = 30 DB; NUMBER OF DATA SAMPLES = 64; RECTANGULAR ENVELOPE; ACTUAL DELAY = 64; AR ORDERS 4, 8 & 16

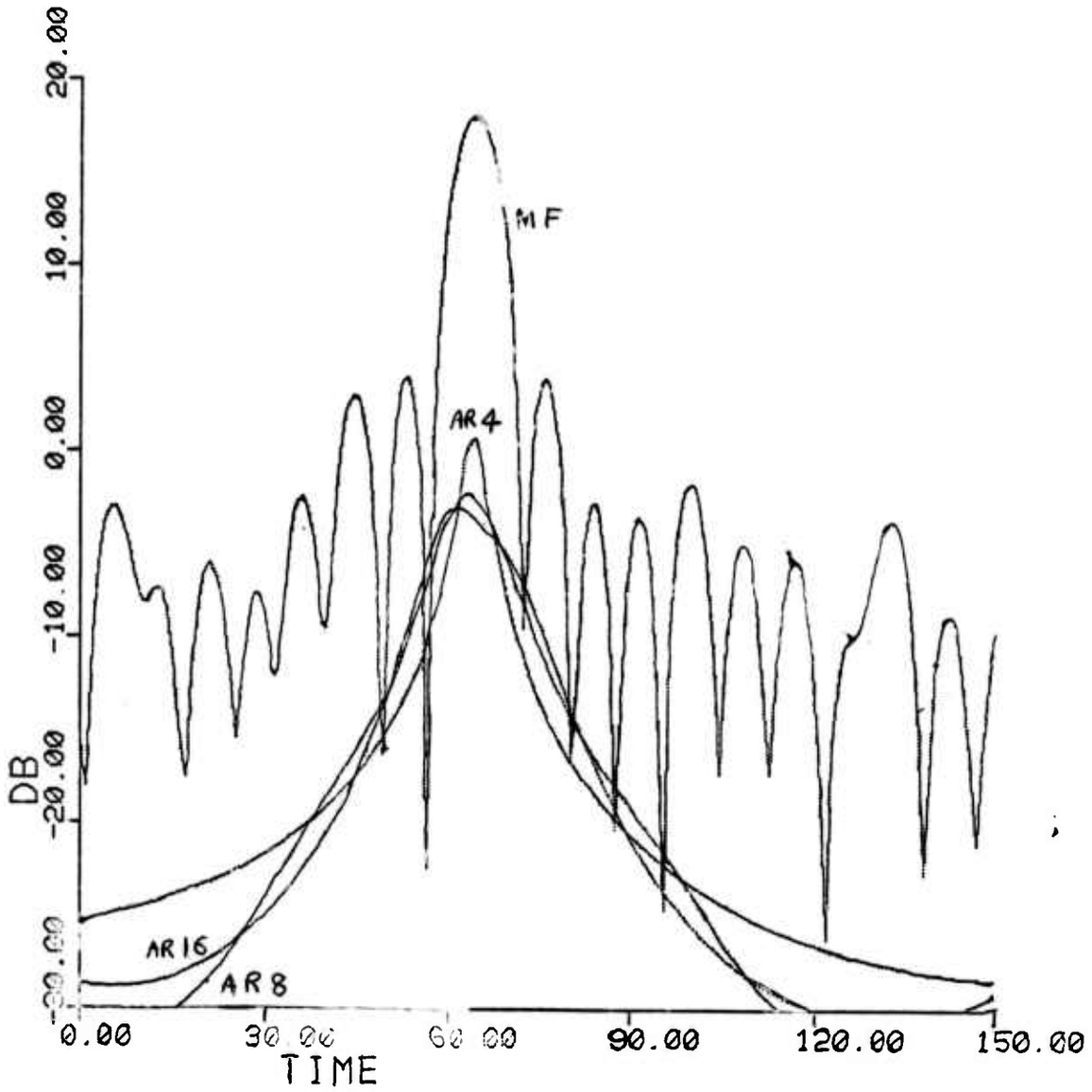


FIG. 8(B)
SAME AS FIG. 7(B) EXCEPT SNR = 20 DB.

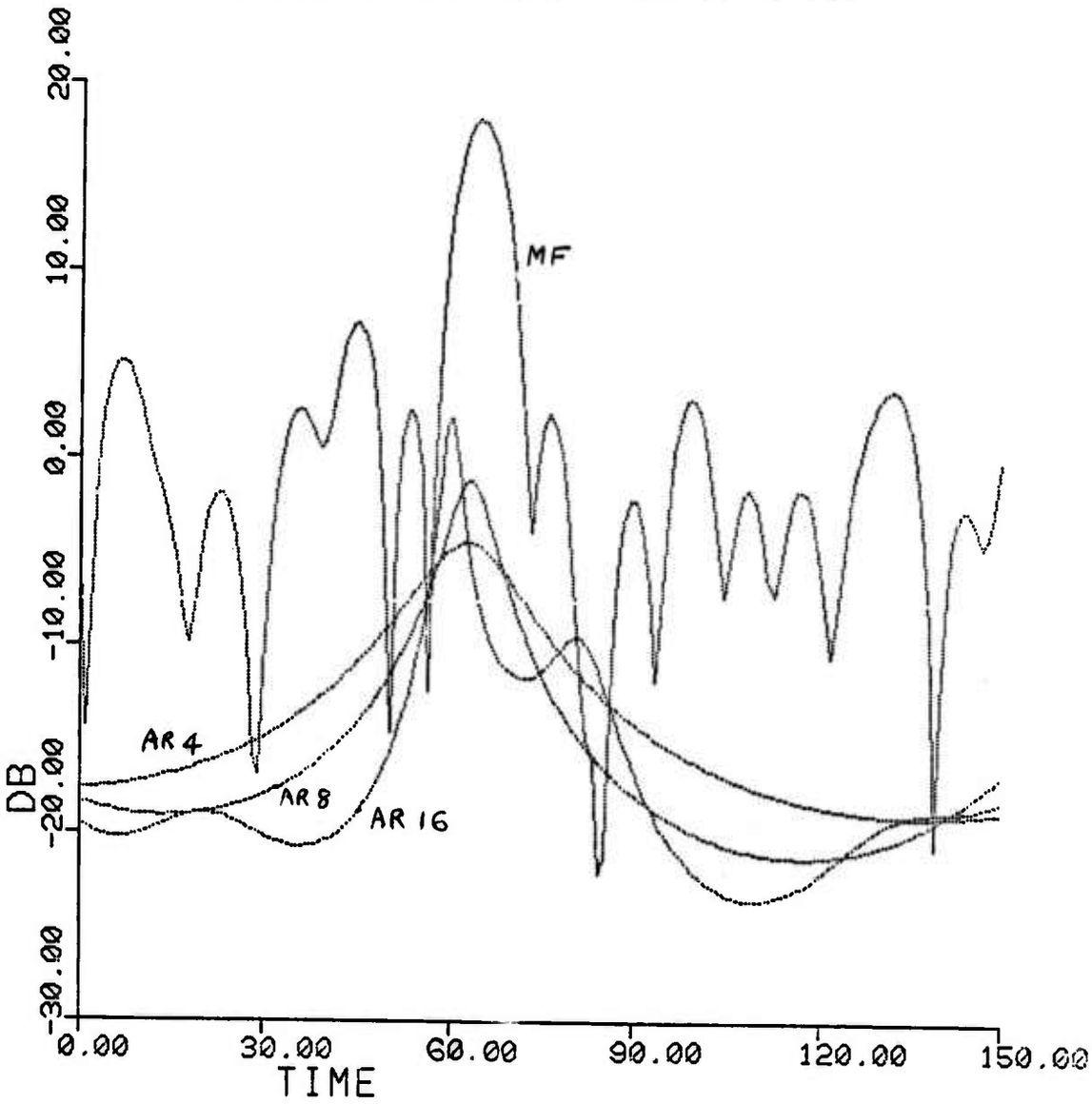


FIG. 9(B)
COMPARISON OF BURG ALGORITHM & MATCHED FILTER FOR DELAY ESTIMATION; SNR = 30 DB; GAUSSIAN ENVELOPE; NUMBER OF DATA SAMPLES = 128; ACTUAL DELAY = 64; AR ORDERS 4, 8 & 16.

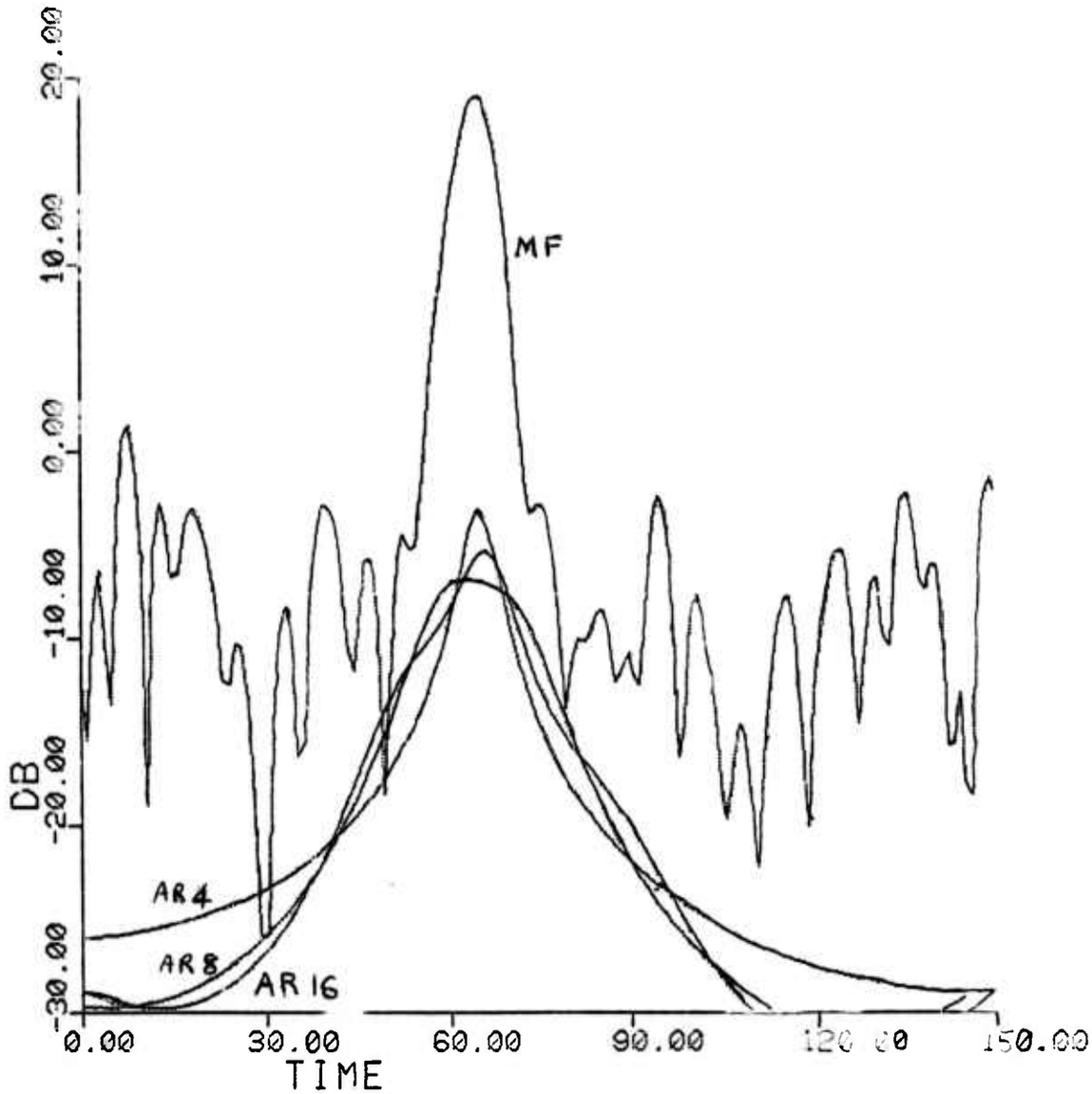


FIG. 10(B)
SAME FIG. 9(B) EXCEPT SNR = 20 DB.

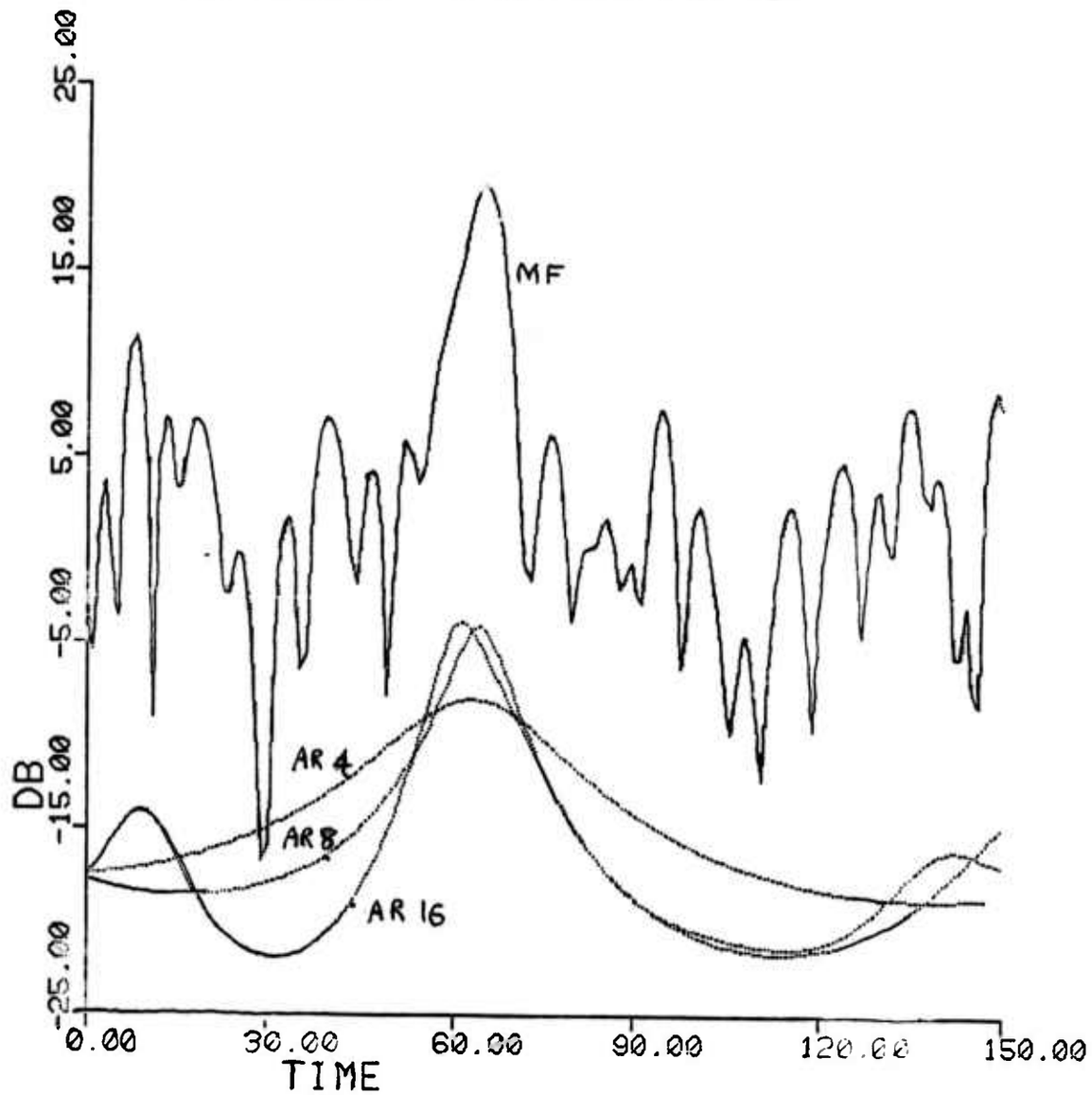


FIG. 11(B)
SAME AS FIG. 9(B) EXCEPT ENVELOPE IS TRIANGULAR

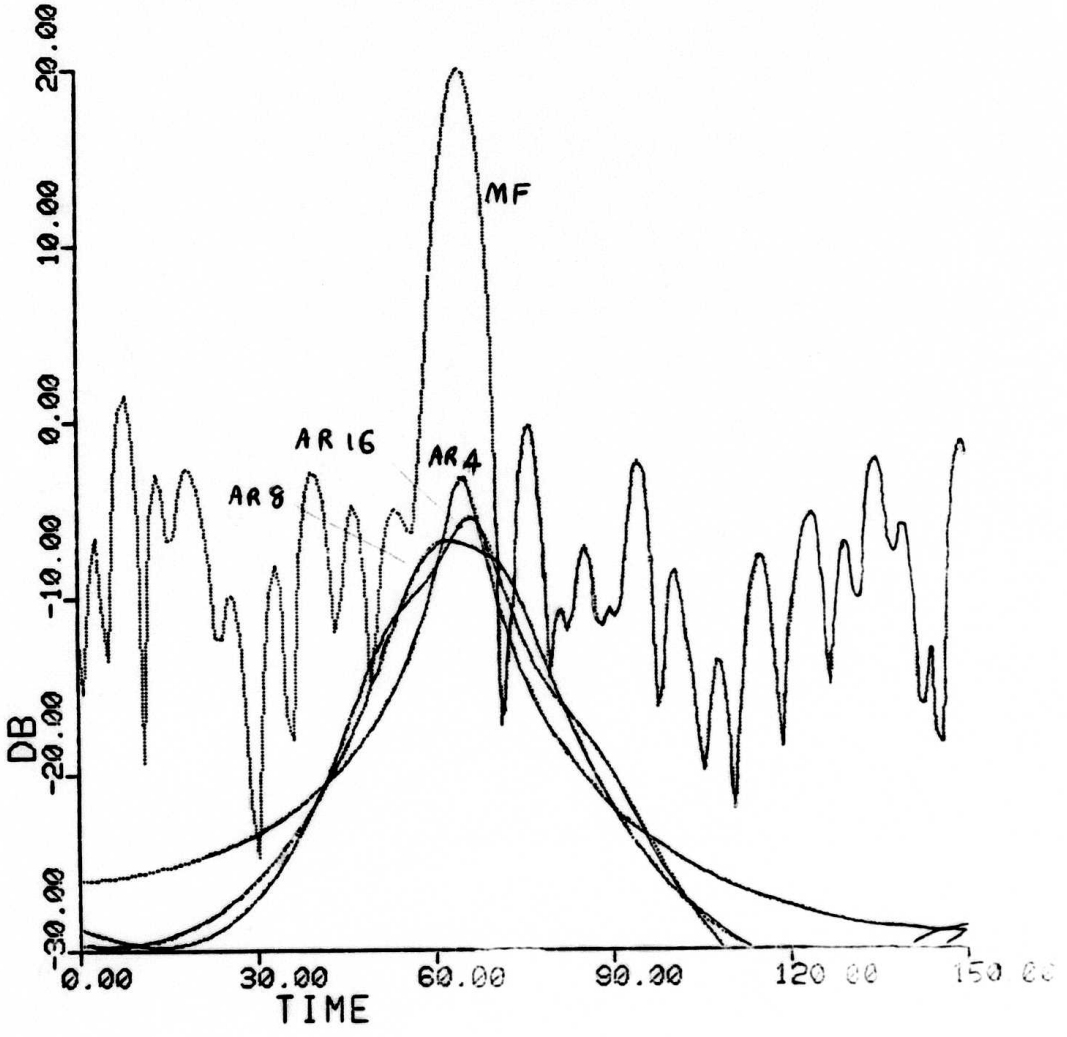


FIG. 12(B)
COMPARISON OF BURG ALGORITHM & MATCHED FILTER FOR DELAY ESTIMATION; SNR = 30 DB; NUMBER OF DATA SAMPLES = 32; GAUSSIAN ENVELOPE; ACTUAL DELAY = 64; AR ORDERS 4, 8 & 16.

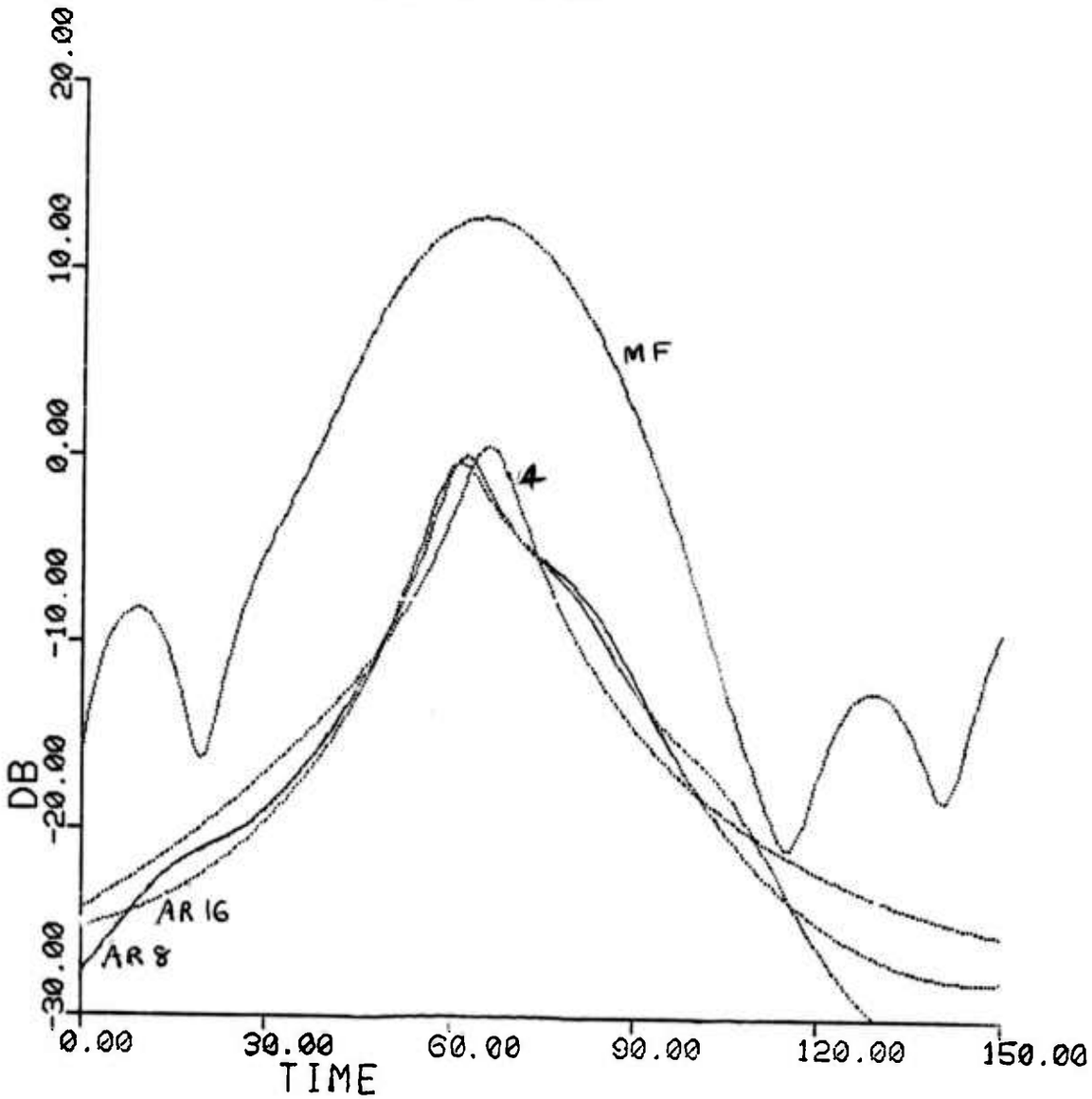


FIG. 13(B)
SAME AS FIG. 12(B), EXCEPT SNR = 20 DB.

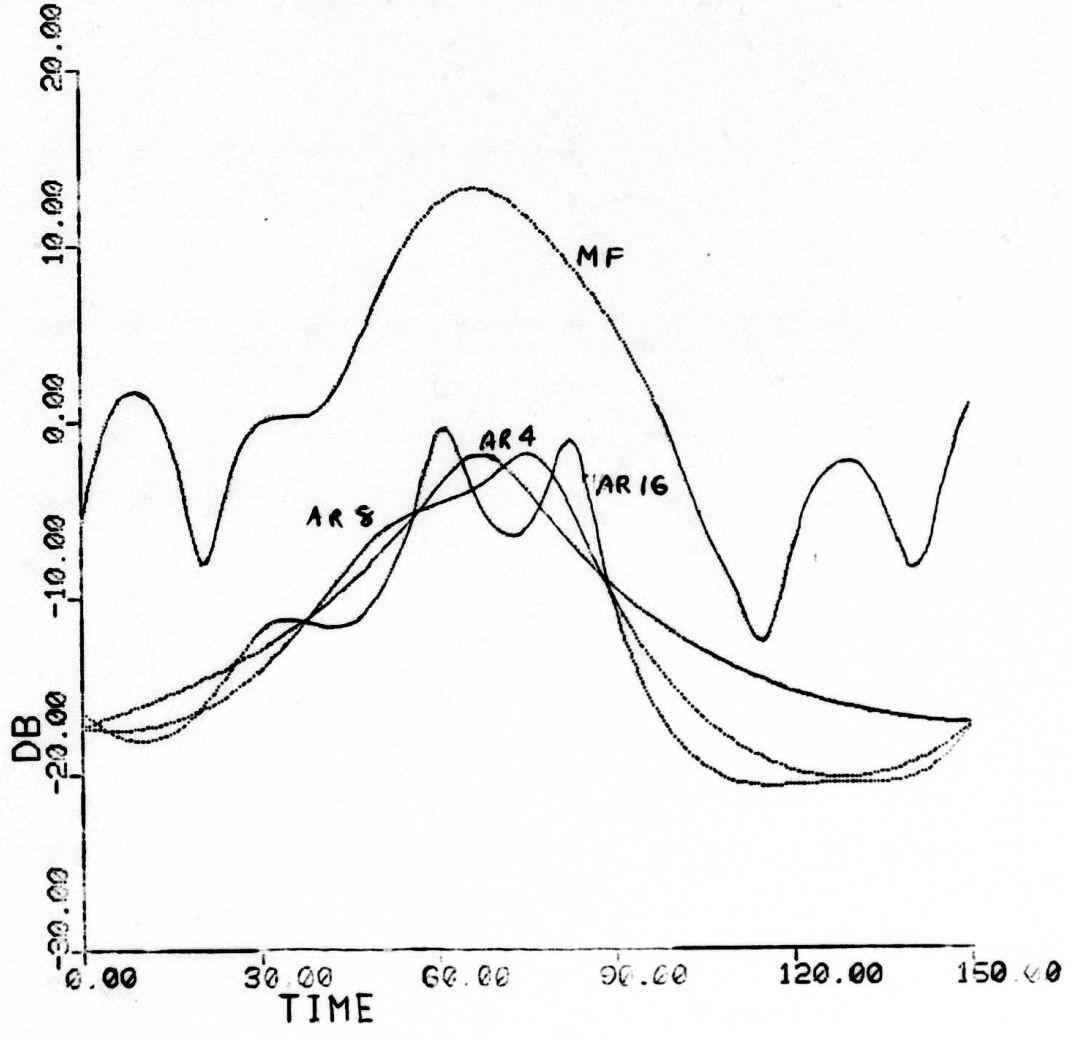


FIG. 14(B)
SAME AS FIG. 1(B) EXCEPT ACTUAL DELAY = 125

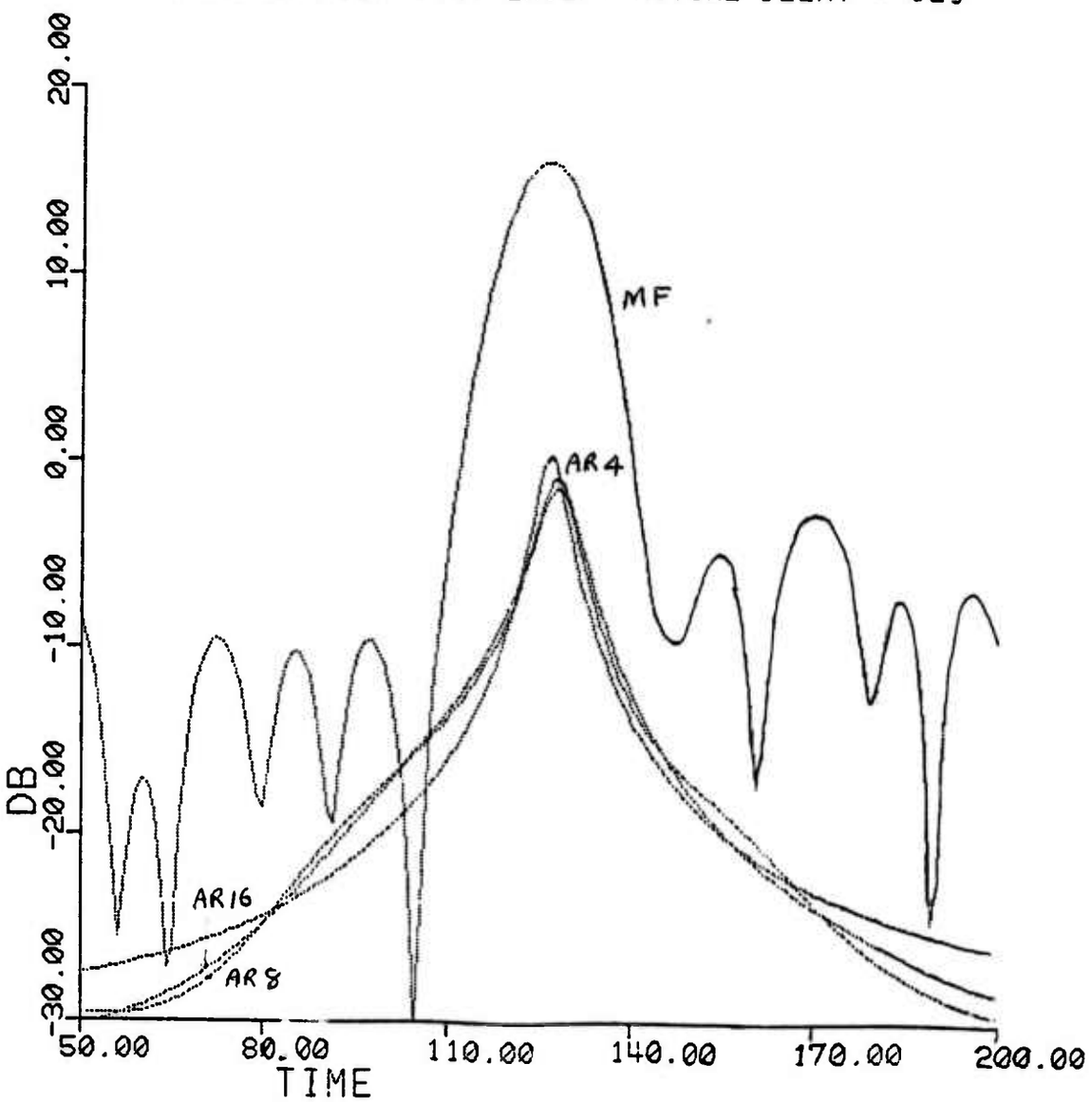


FIG. 15(B)
COMPARISON OF BURG ALGORITHM & MATCHED FILTER FOR DELAY ESTIMATION; SNR = 30 DB; NUMBER OF DATA SAMPLES = 128; ACTUAL DELAY = 125; GAUSSIAN ENVELOPE; AR ORDERS 4, 8 & 16

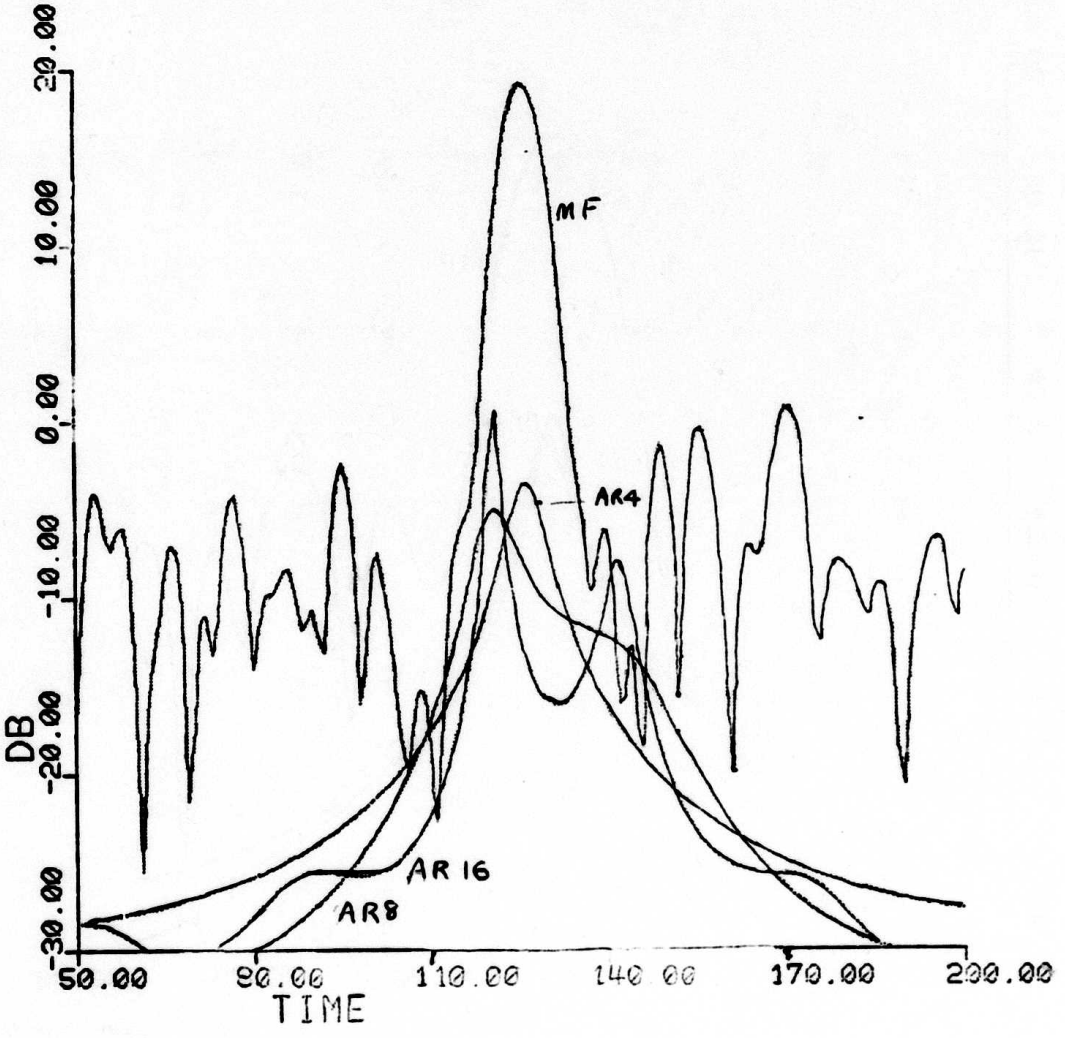


FIG. 16(B)
SAME AS FIG. 15(B) BUT NUMBER OF SAMPLES = 32

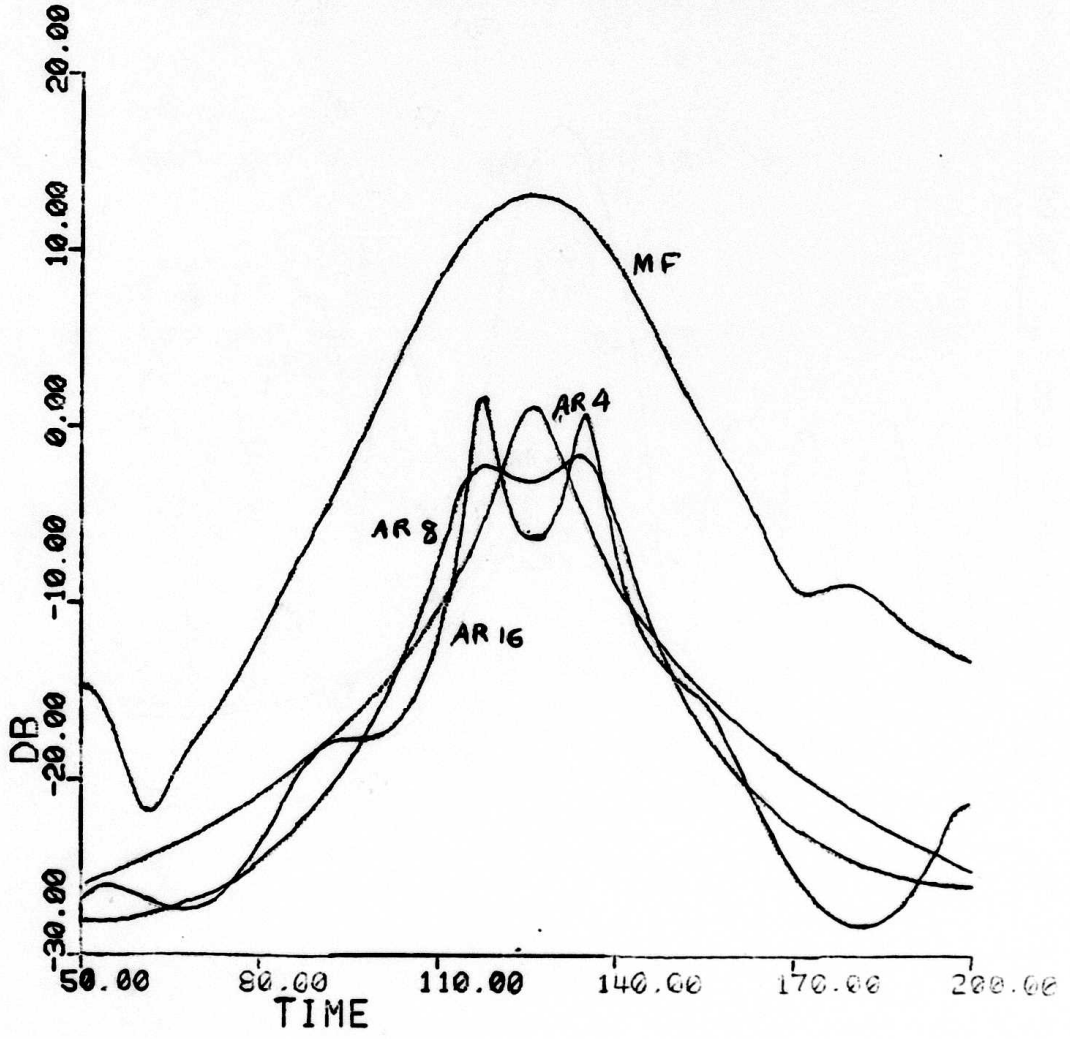


FIG. 17(B)

COMPARISON OF DELAY ESTIMATES WITH BURG ALGORITHM & MATCHED FILTER; SNR = 20 DB; NUMBER OF DATA SAMPLES = 64; GAUSSIAN ENVELOPE; ACTUAL DELAY = 125; AR ORDERS 4, 8 & 16

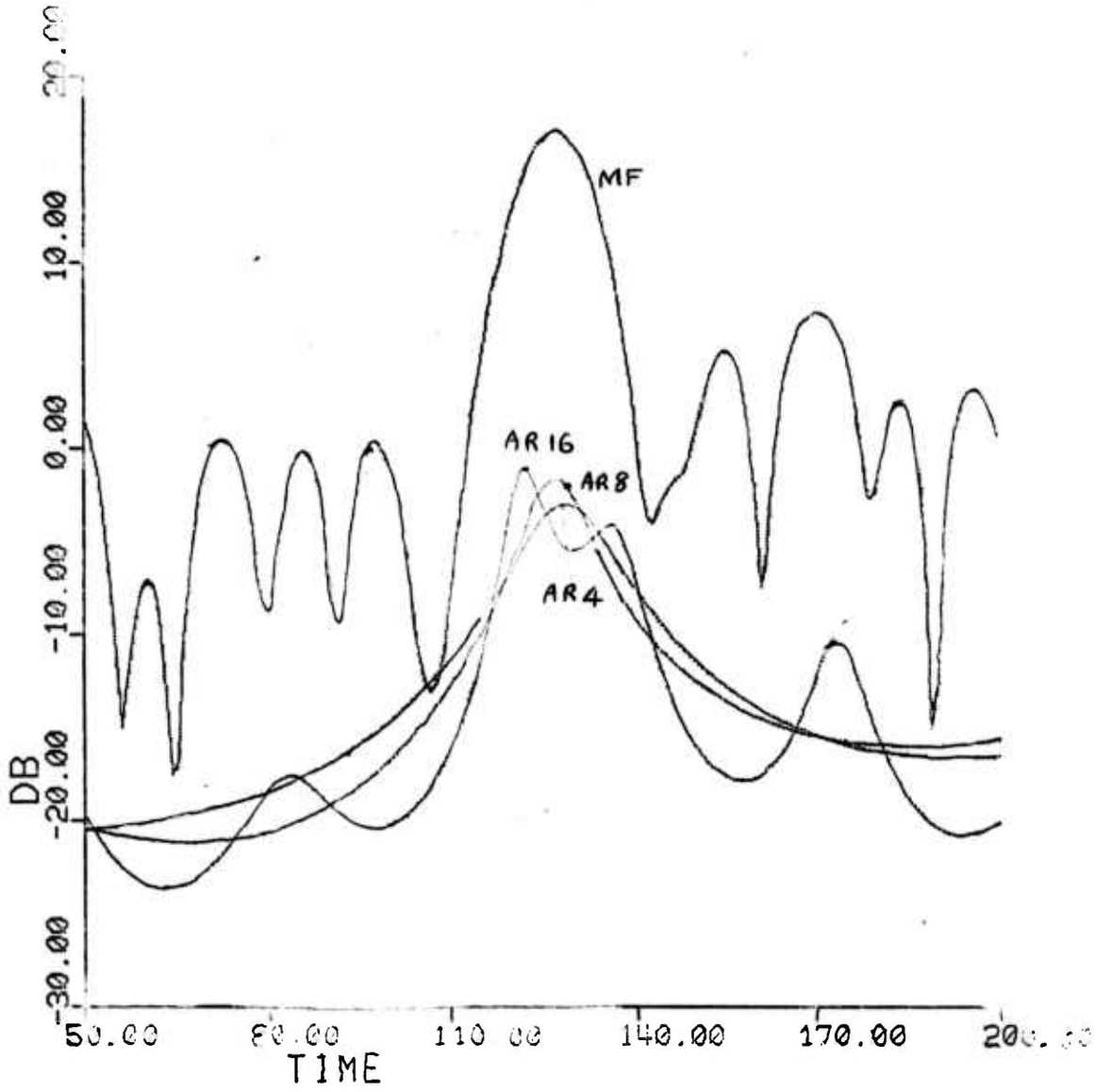


FIG. 18(B)
SAME AS FIG. 17(B) EXCEPT NUMBER OF SAMPLES = 128

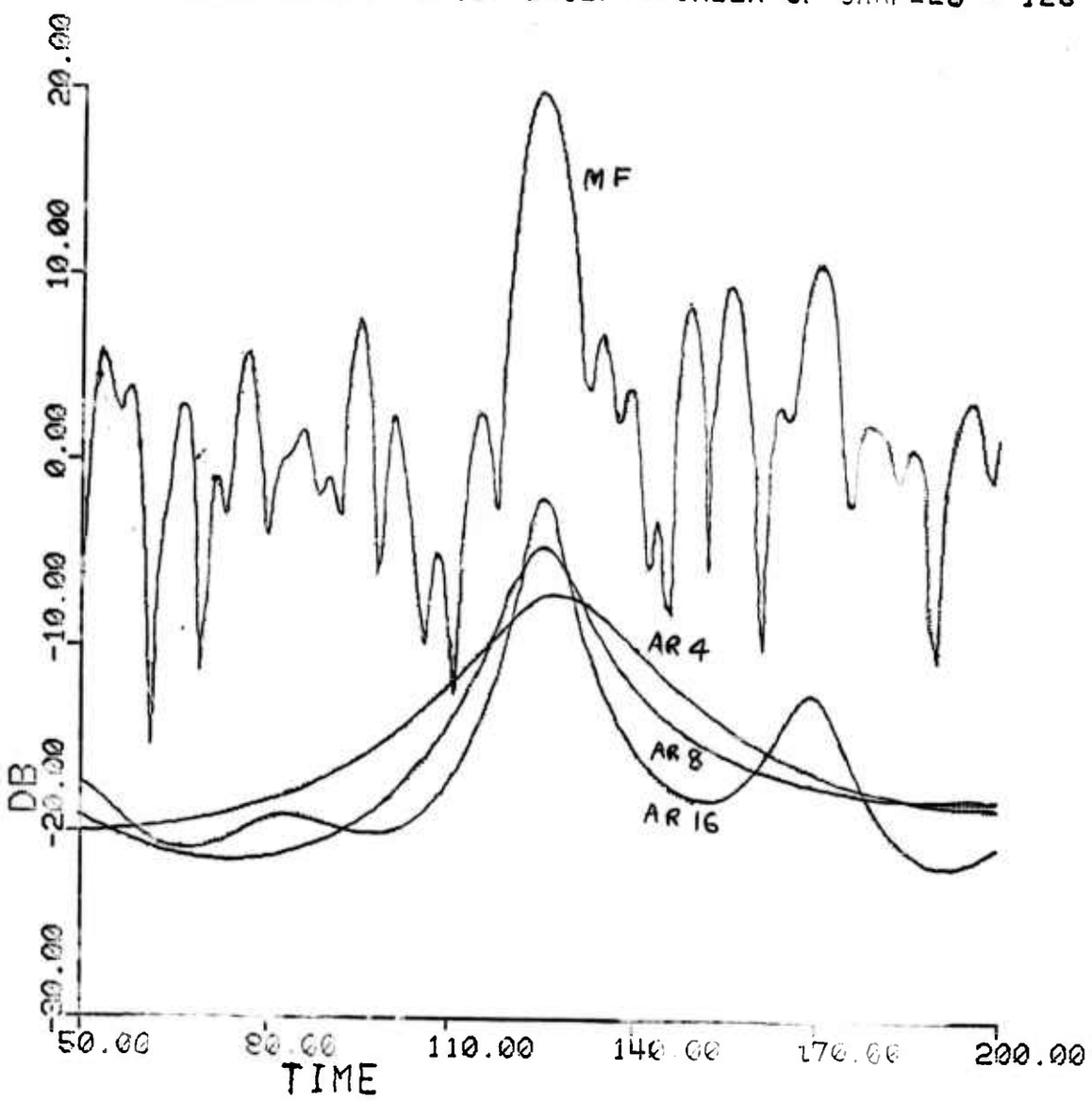


FIG. 19(B)
SAME AS FIG. 17(B) BUT NUMBER OF SAMPLES = 32.

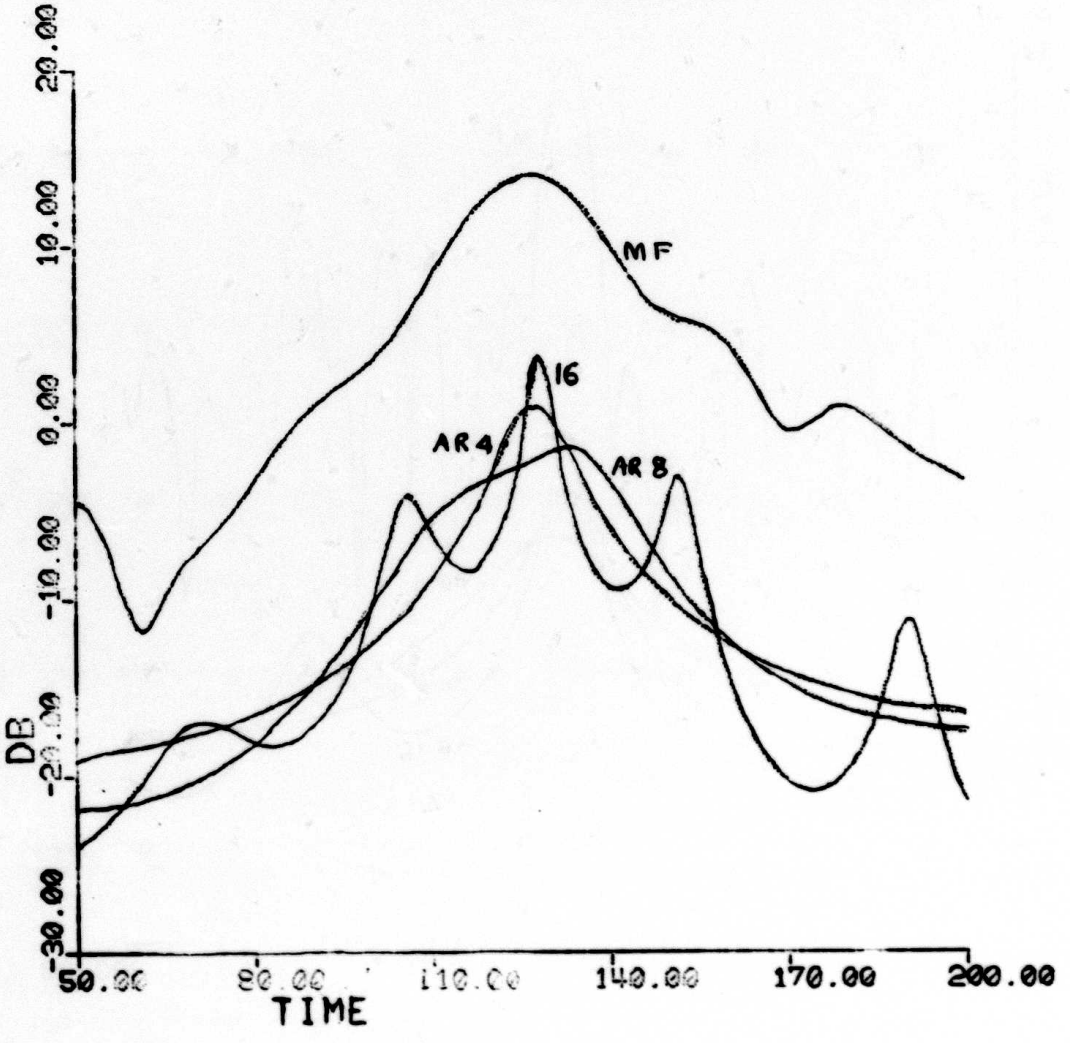


FIG. 20(B)
COMPARISON OF BURG ALGORITHM & MATCHED FILTER FOR DELAY ESTIMATION; SNR = 30 DB; NUMBER OF DATA SAMPLES = 64; GAUSSIAN ENVELOPE; ACTUAL DELAY = 25; AR ORDERS 4, 8 & 16

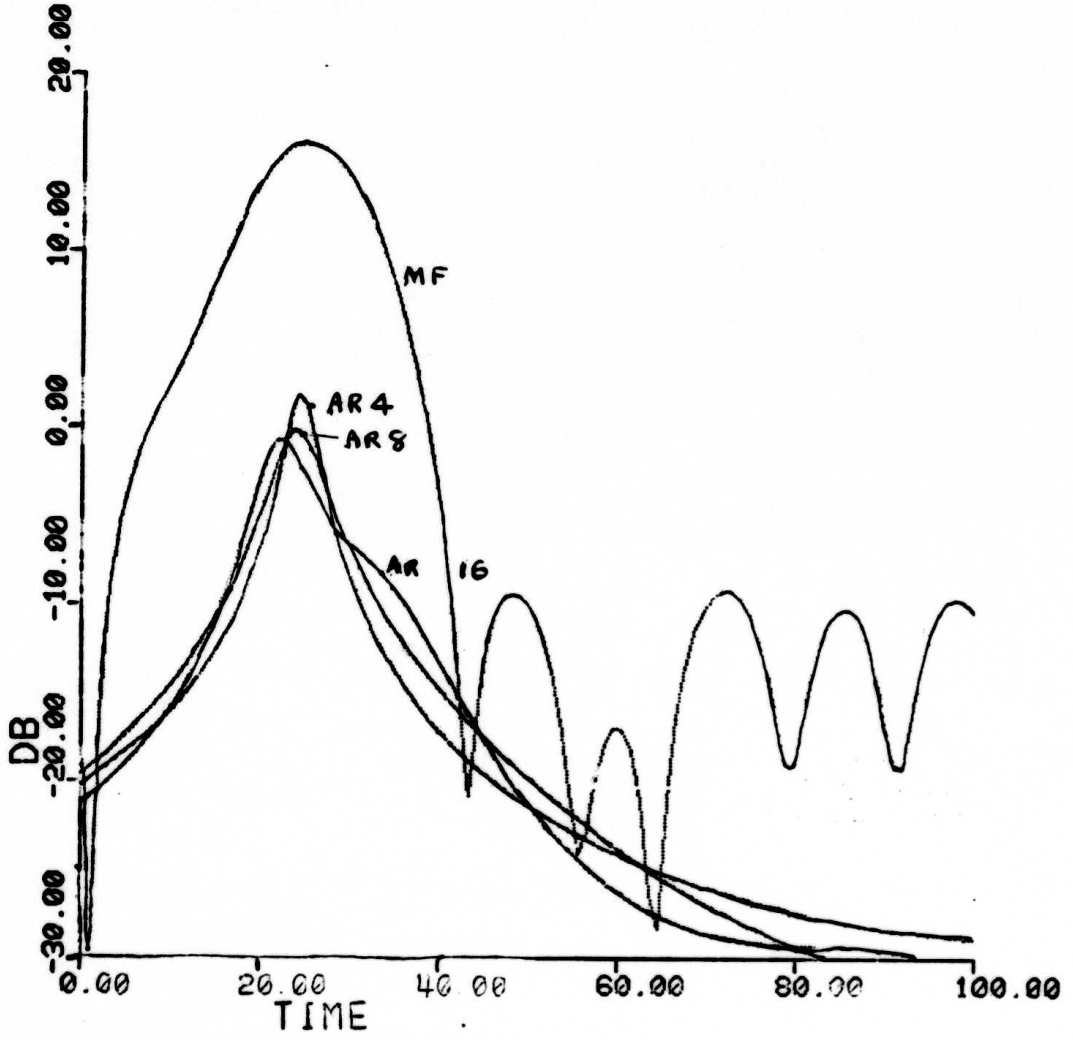


FIG. 21(B)
SAME AS FIG. 20(B) BUT NUMBER OF SAMPLES = 128

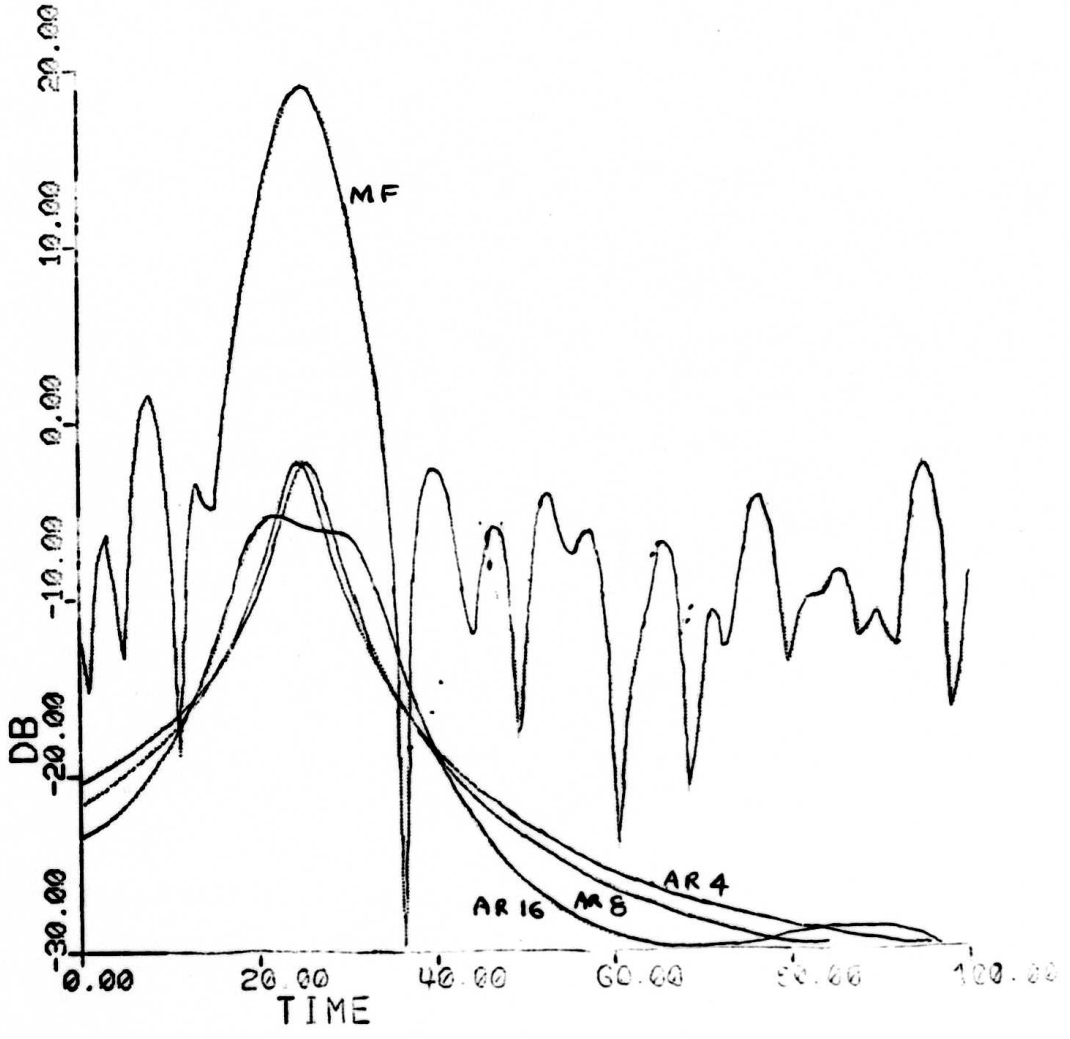


FIG. 22(B)
SAME AS FIG. 20(B) BUT NUMBER OF SAMPLES = 32

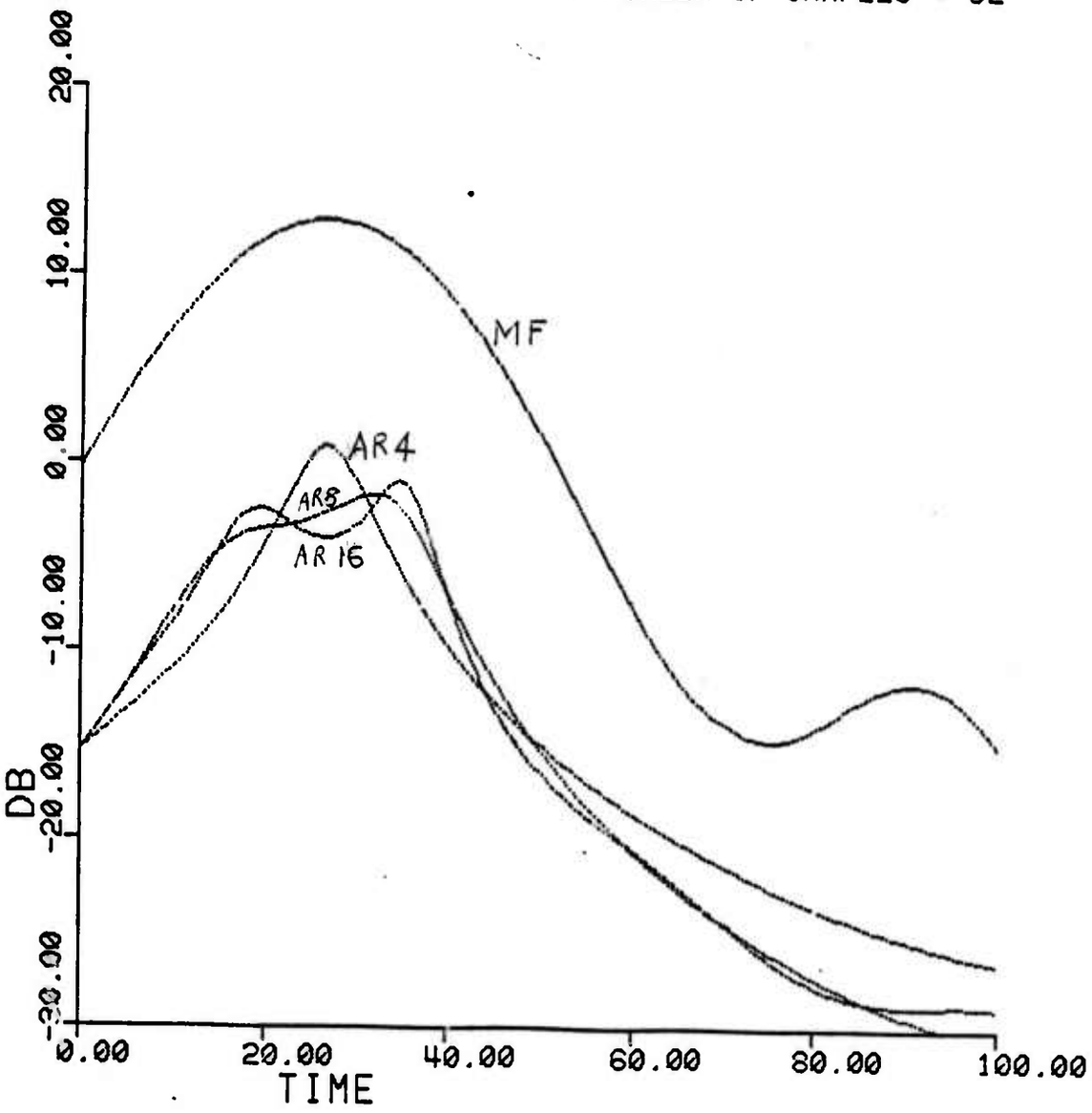


FIG. 23(B)
COMPARISON OF BURG ALGORITHM & MATCHED FILTER FOR DELAY ESTIMATION: SNR = 20 DB; NUMBER OF SAMPLES = 64; GAUSSIAN ENVELOPE; ACTUAL DELAY = 25; AR ORDERS 4, 8 & 16

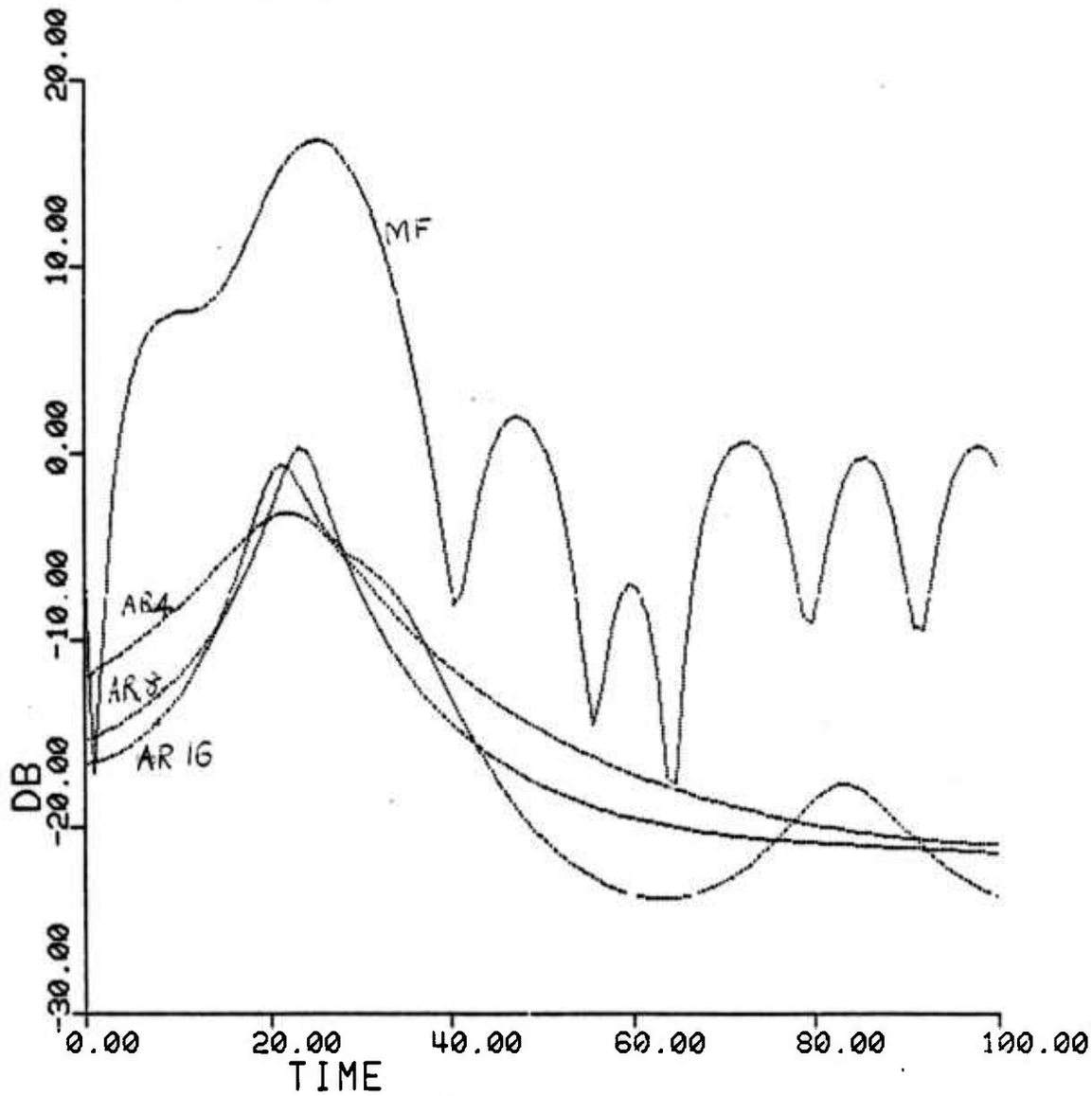


FIG. 24(B)
DELAY ESTIMATION WITH BURG ALGORITHM AND MATCHED FILTER; DATA : SNR = 30 DB; GAUSSIAN ENVELOPE; NUMER OF SAMPLES = 64; ACTUAL DELAYS OF EQUAL STRENGTH AT 64 & 100; AR ORDERS 4, 8 & 16

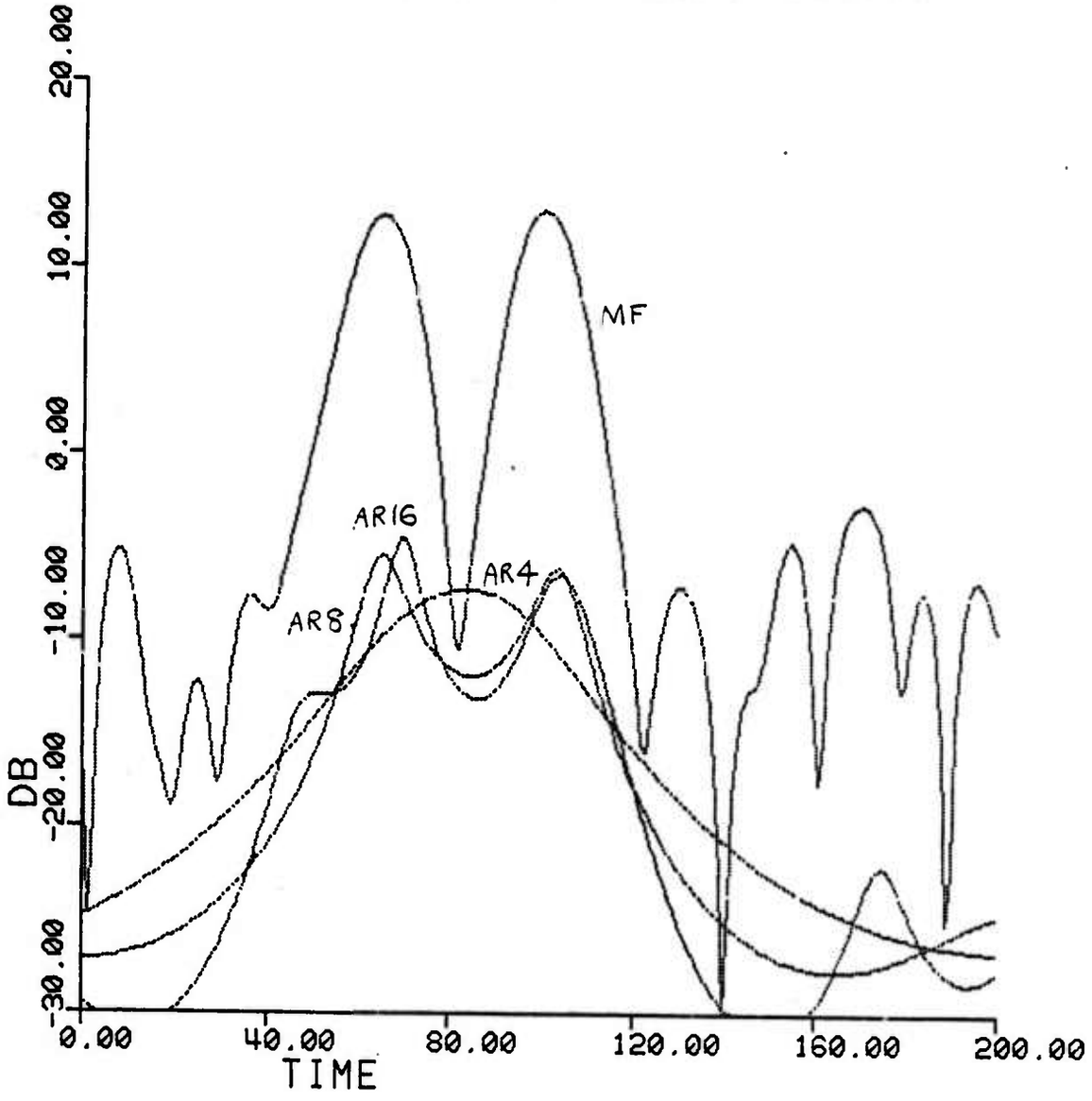


FIG. 25(B)
DELAY ESTIMATION WITH BURG ALGORITHM & MATCHED FILTER; DATA : SNR = 30 DB; GAUSSIAN ENVELOPE; NUMBER OF SAMPLES = 64; ACTUAL DELAYS OF EQUAL AMPLITUDE AT 64 & 80; AR ORDERS 4, 8 & 16.

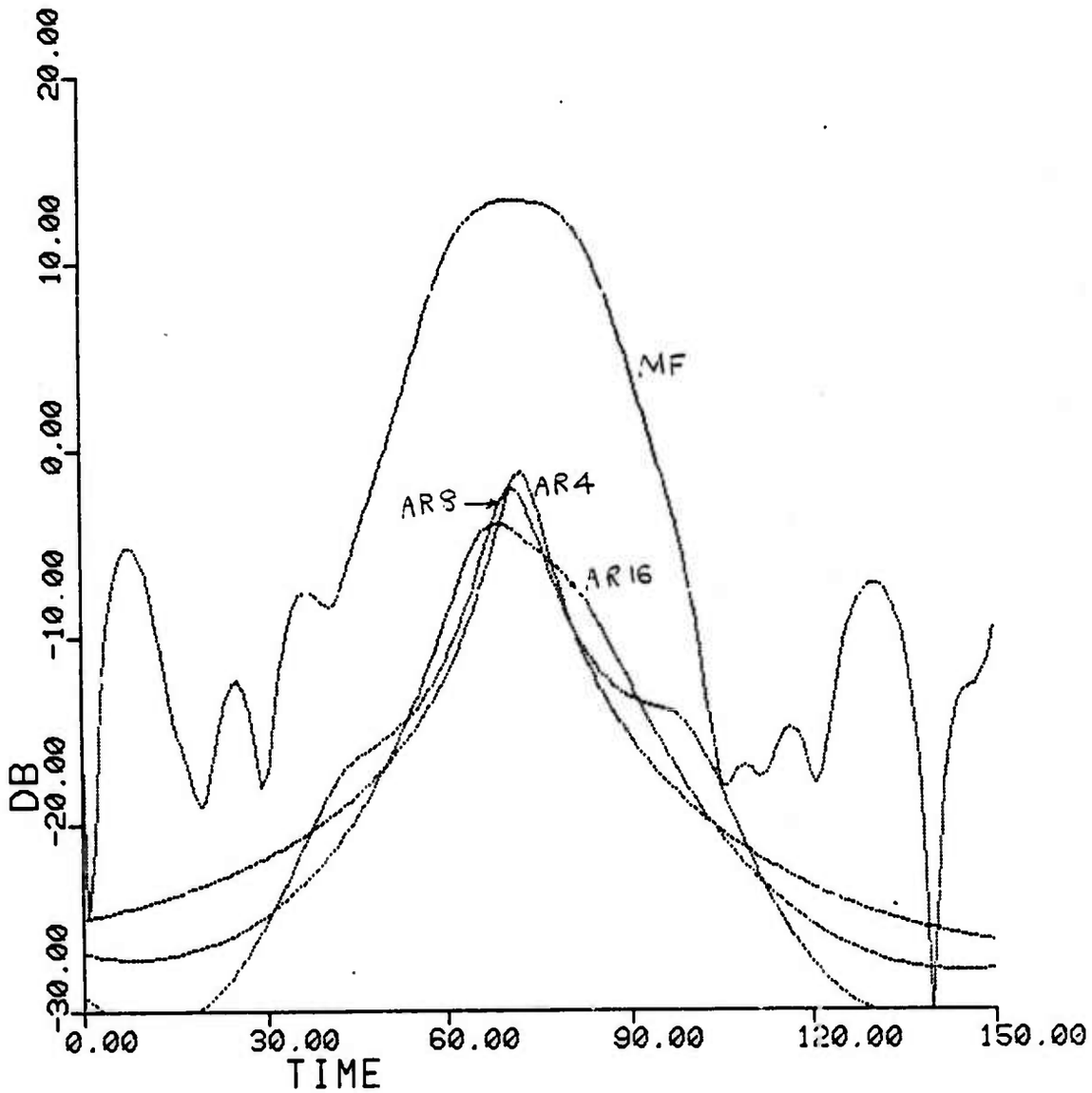


FIG. 26(B)
DELAY ESTIMATION WITH MATCHED FILTER AND BURG ALGORITHM; DATA : SNR = 30 DB; GAUSSIAN ENVELOPE NUMBER OF SAMPLES = 64; ACTUAL DELAYS OF EQUAL AMPLITUDE AT 64 & 84; AR ORDERS 8, 16 & 32

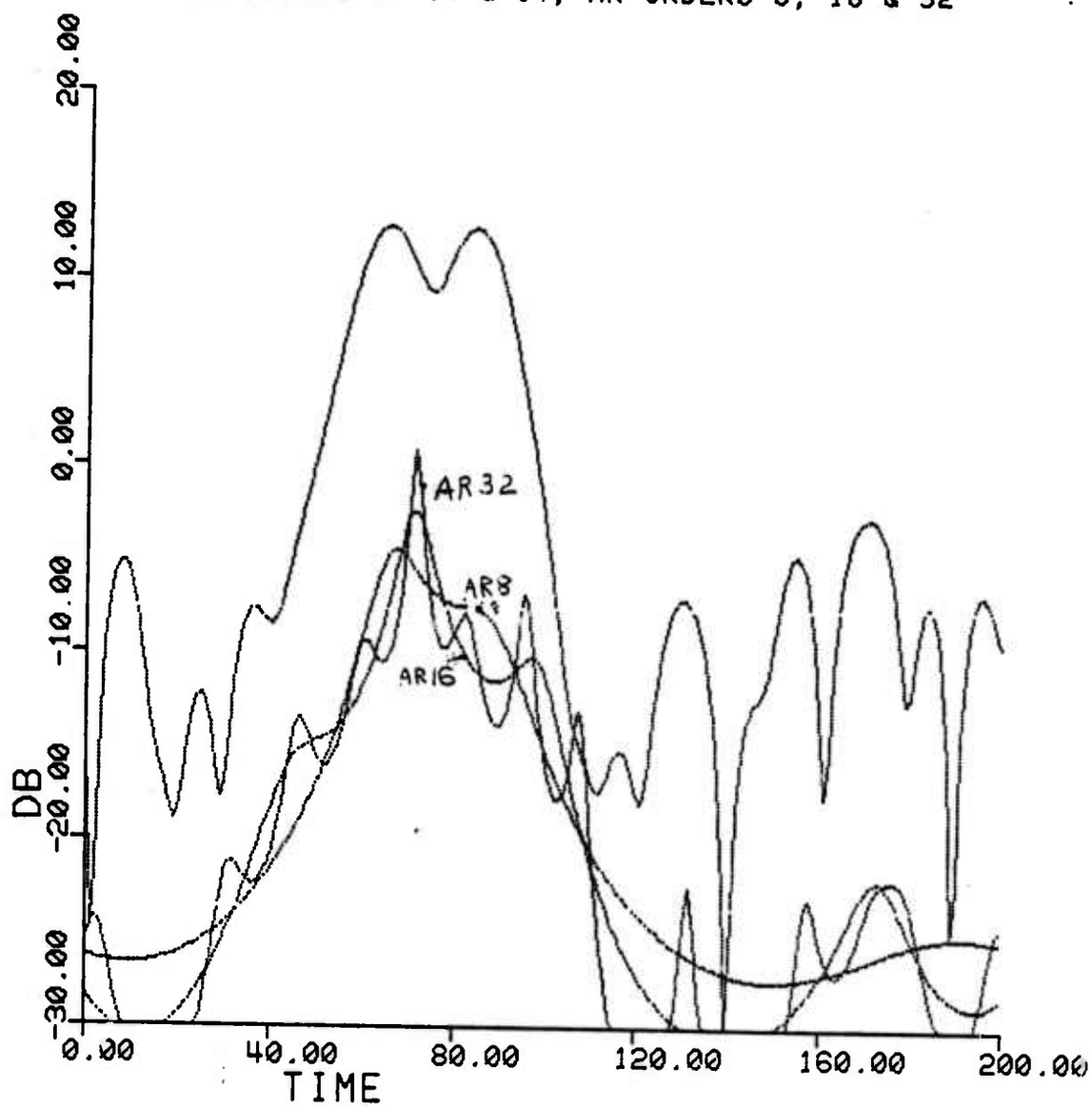


FIG. 27(B)
DELAY ESTIMATION WITH BURG ALGORITHM & MATCHED FILTER; DATA : SNR = 30 DB; GAUSSIAN ENVELOPE; ACTUAL DELAYS OF EQUAL AMPLITUDE AT 64 & 84; NUMBER OF SAMPLES =128; AR ORDERS 8 & 16.

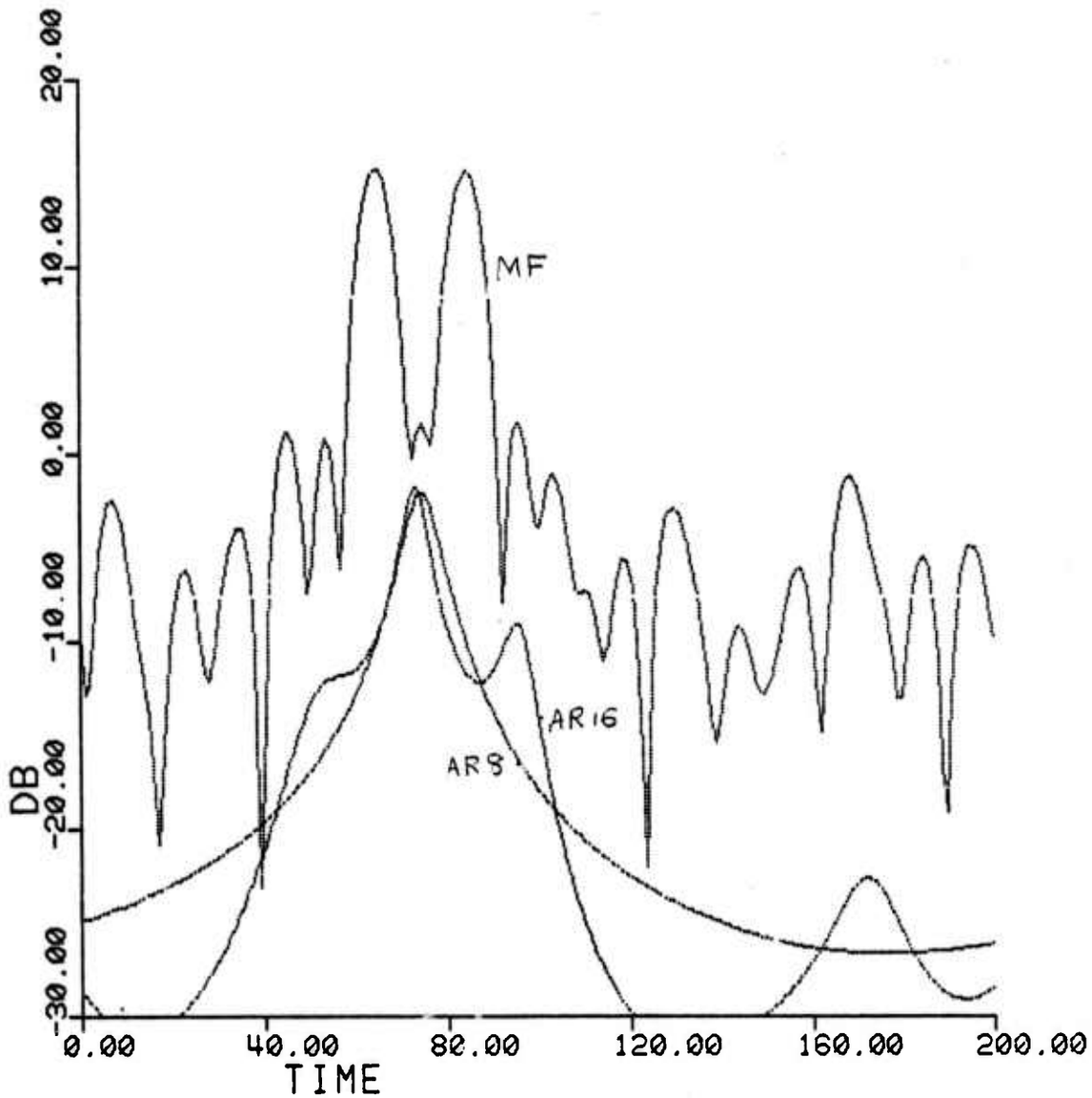


FIG. 28(B)
DELAY ESTIMATION WITH BURG ALGORITHM & MATCHED FILTER; SNR = 30 DB; GAUSSIAN ENVELOPE; NUMBER OF DATA SAMPLES = 64; ACTUAL DELAYS AT 25 & 100; AR ORDERS 4, 8 & 16.

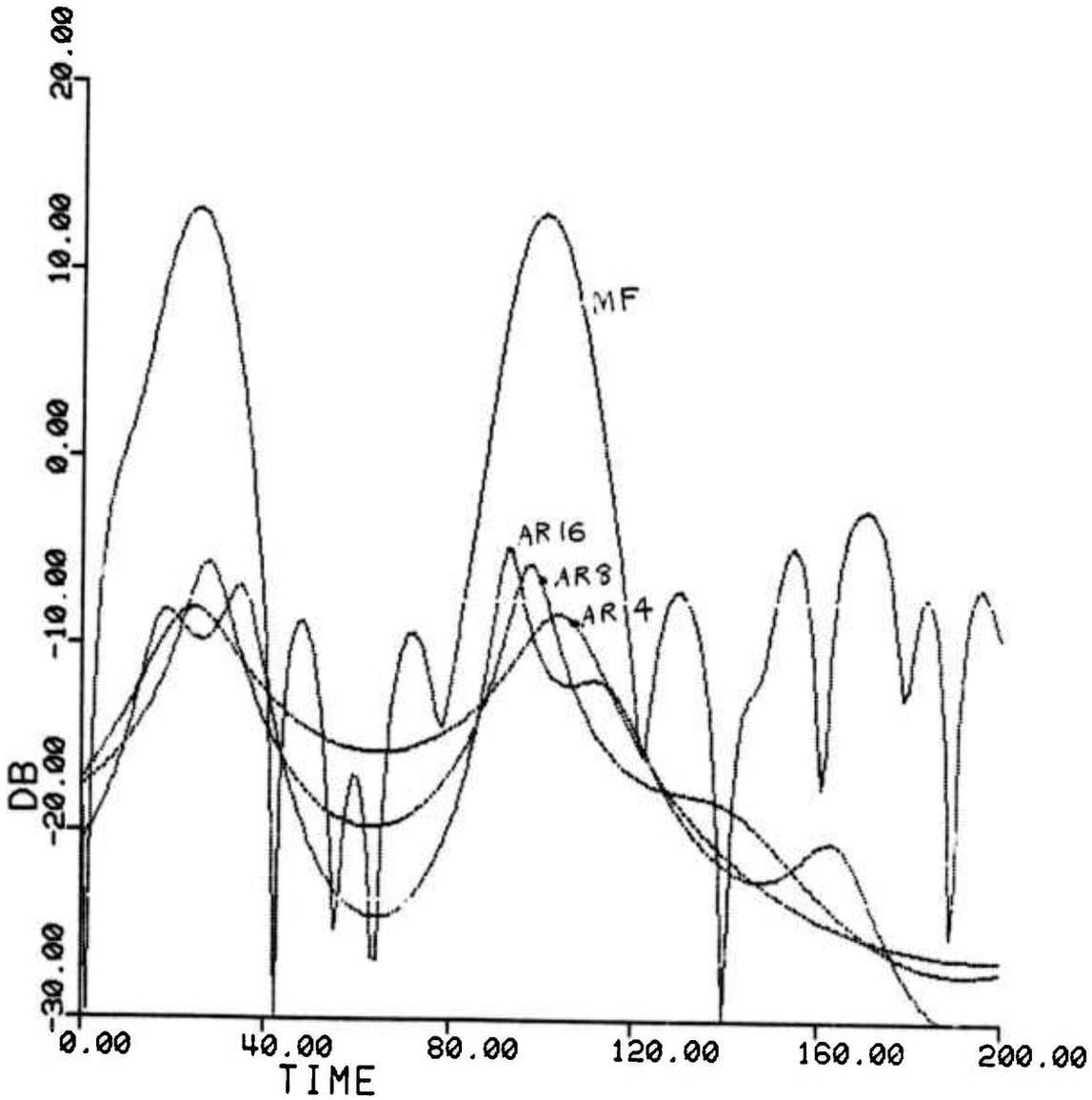


FIG. 29(B)
DELAY ESTIMATION WITH MATCHED FILTER & BURG ALGORITHM; SNR = 30 DB; GAUSSIAN ENVELOPE; NUMBER OF SAMPLES = 64; TWO DELAYS OF EQUAL AMPLITUDE AT 25 & 35; AR ORDERS = 4, 8 & 16.

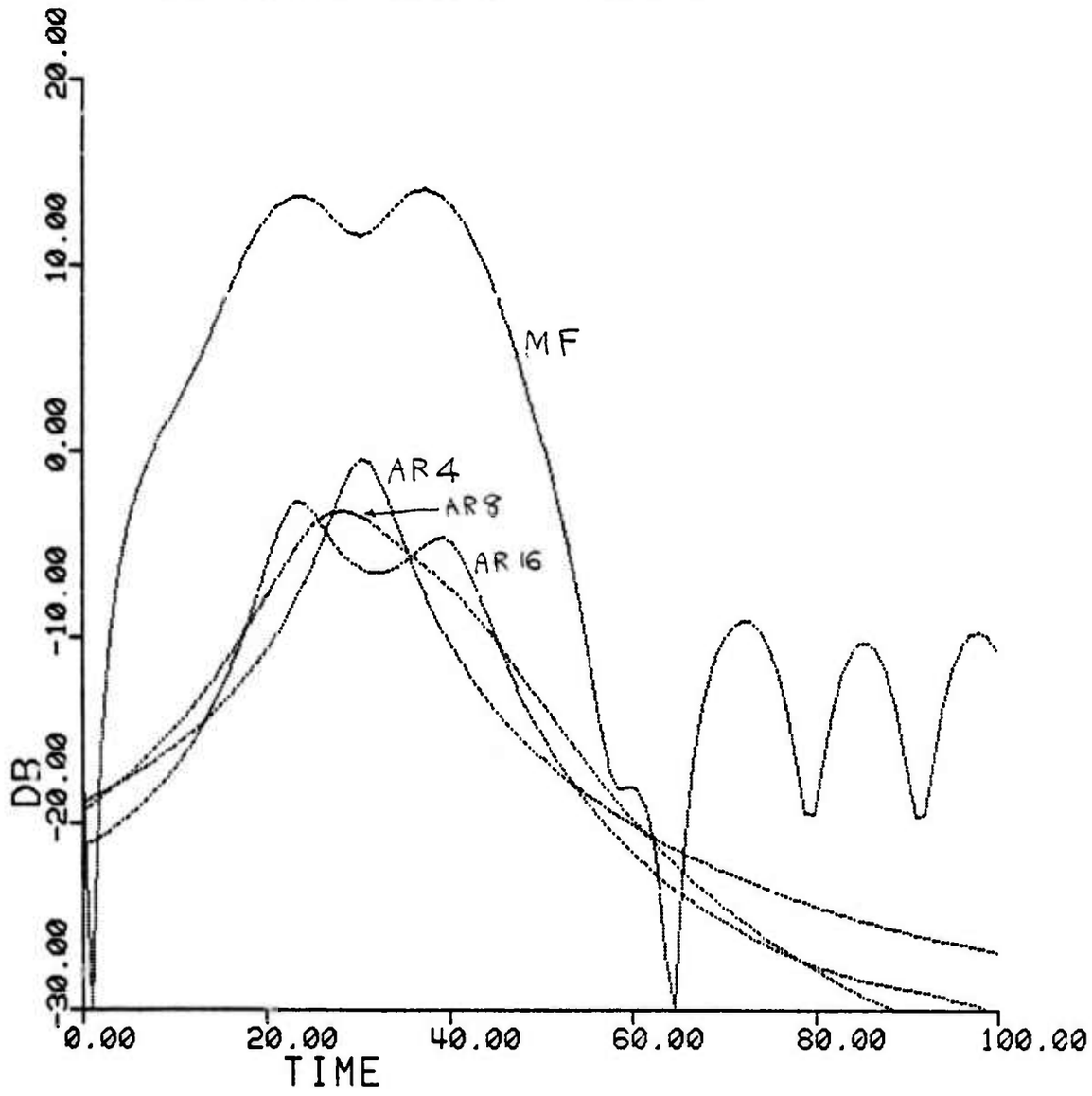


FIG. 30(B)
SAME AS FIG. 29(B) BUT NUMBER OF SAMPLES = 32.

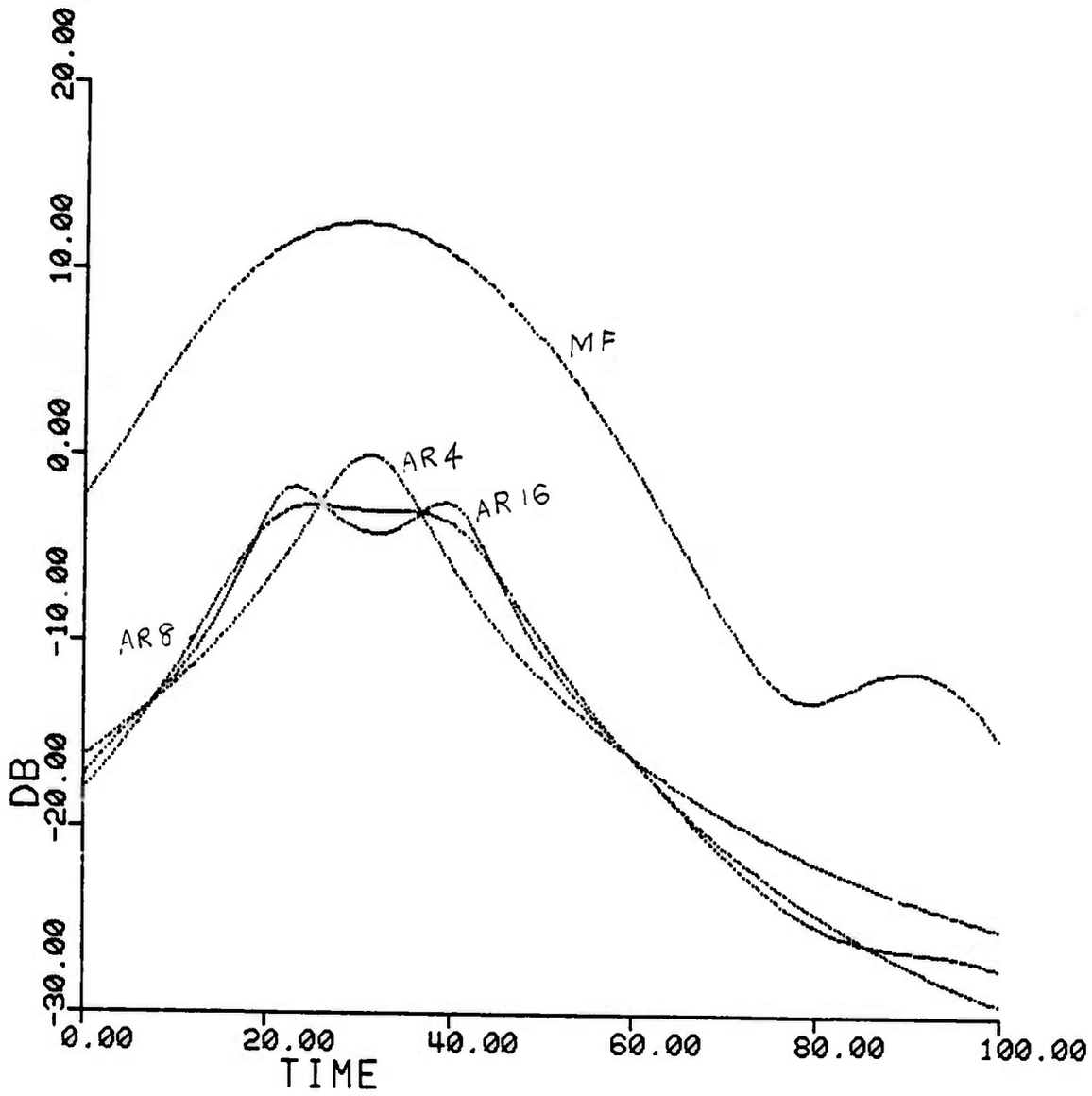


FIG. 31(B)
DELAY ESTIMATION WITH MATCHED FILTER & BURG ALGORITHM; SNR = 20 DB; GAUSSIAN ENVELOPE; NUMBER OF SAMPLES = 64; TWO DELAYS OF EQUAL AMPLITUDE AT 25 & 35; AR ORDERS = 8 & 16

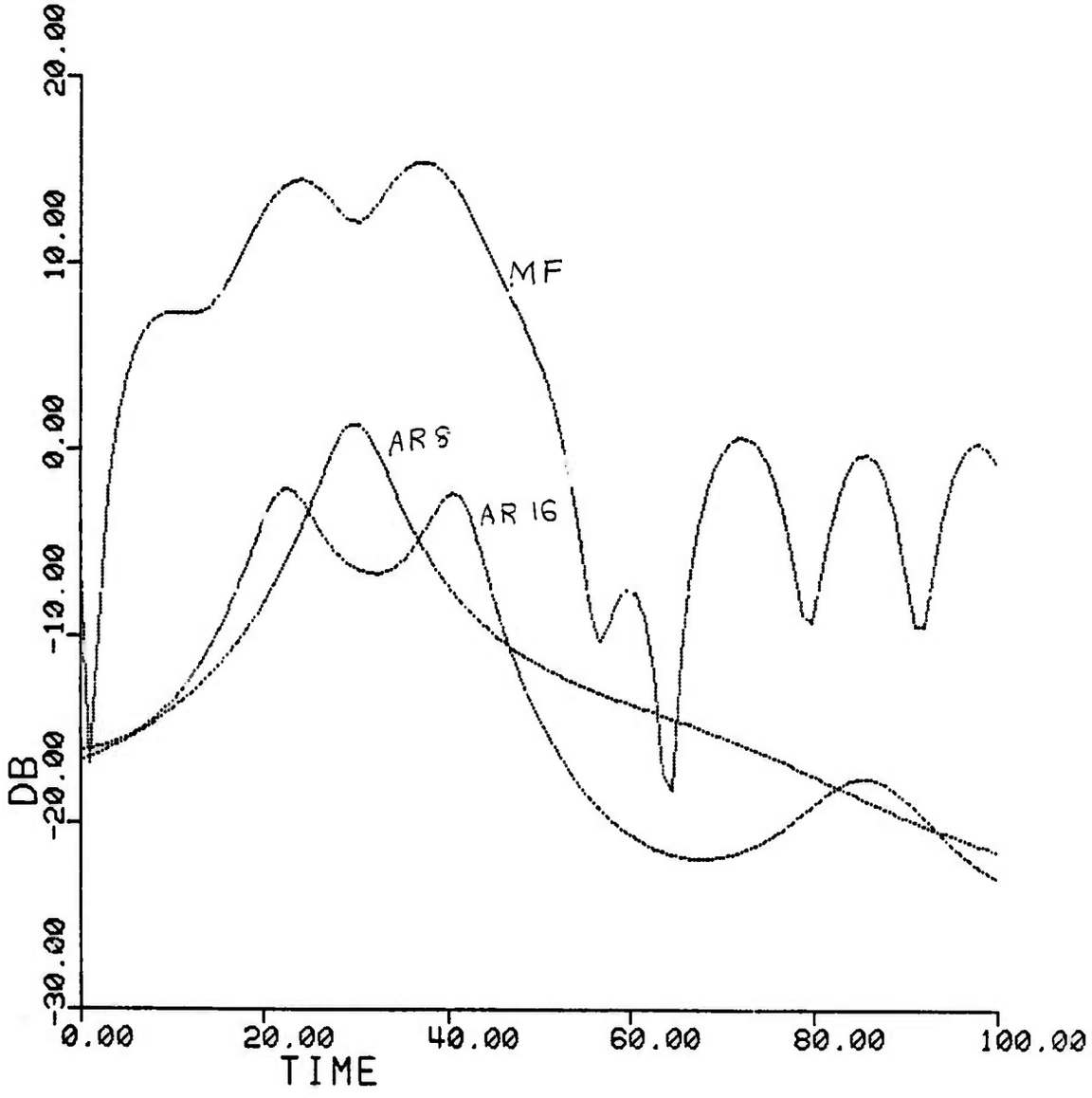


FIG. 32(B)
DELAY ESTIMATION WITH MATCHED FILTER & BURG ALGORITHM; SNR = 20 DB; GAUSSIAN ENVELOPE; NUMBER OF SAMPLES = 32; TWO DELAYS WITH EQUAL AMPLITUDE AT 25 & 35; AR ORDERS 8 & 16.

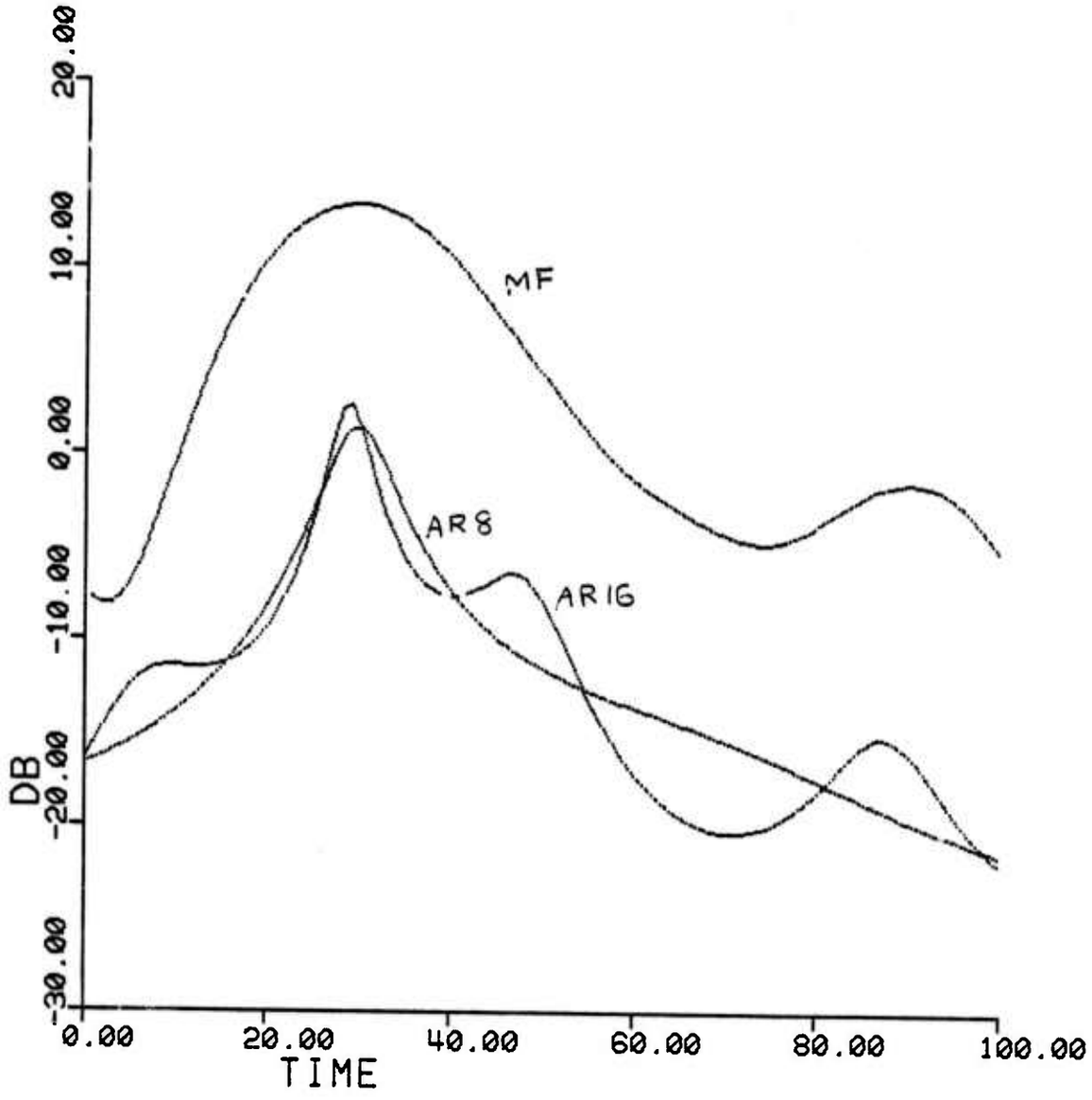


FIG. 33(B)
SAME AS FIG. FIG. 22(B) BUT SNR = 20 DB.

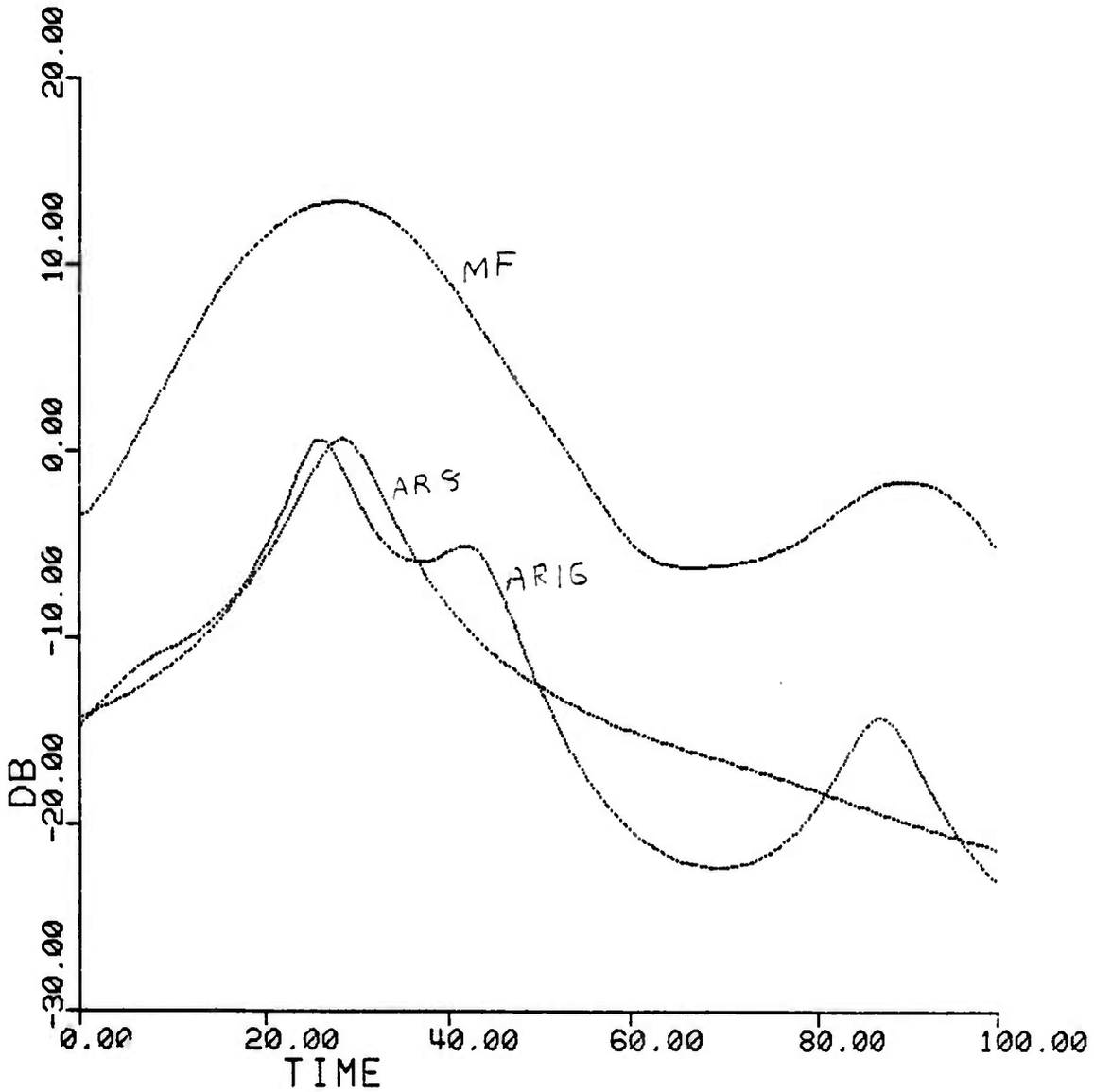


FIG. 34(B)
DELAY ESTIMATION WITH BURG ALGORITHM & MATCHED FILTER; SNR = 20 DB; GAUSSIAN ENVELOPE; NUMBER OF SAMPLES = 64; ONE DELAY AT 25; AR ORDERS = 8 & 16

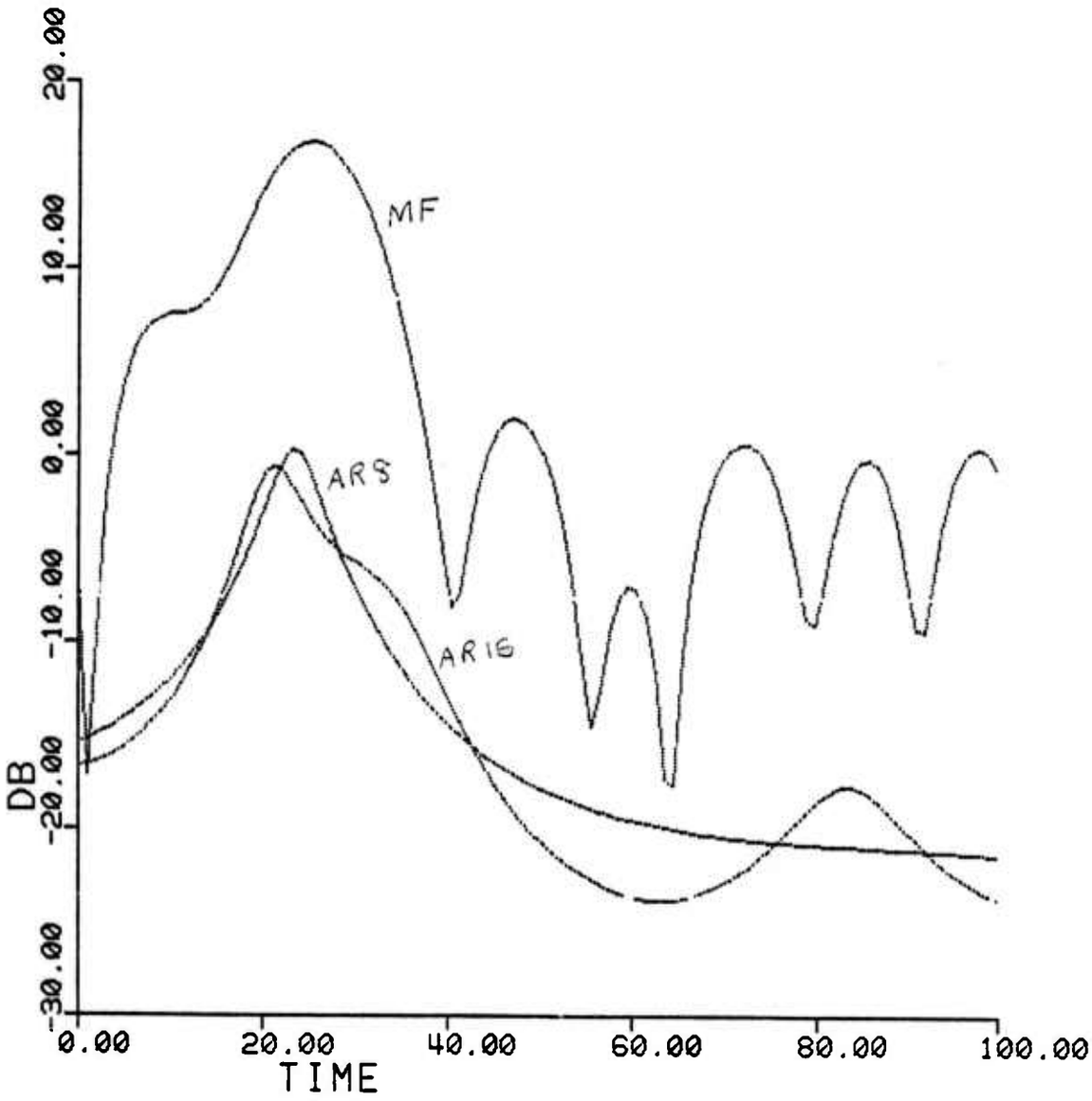


FIG. 1(L)
DELAY ESTIMATION WITH LEVINSON ALGORITHM & MATCHED FILTER. DATA SAME AS IN FIG. 24(B)

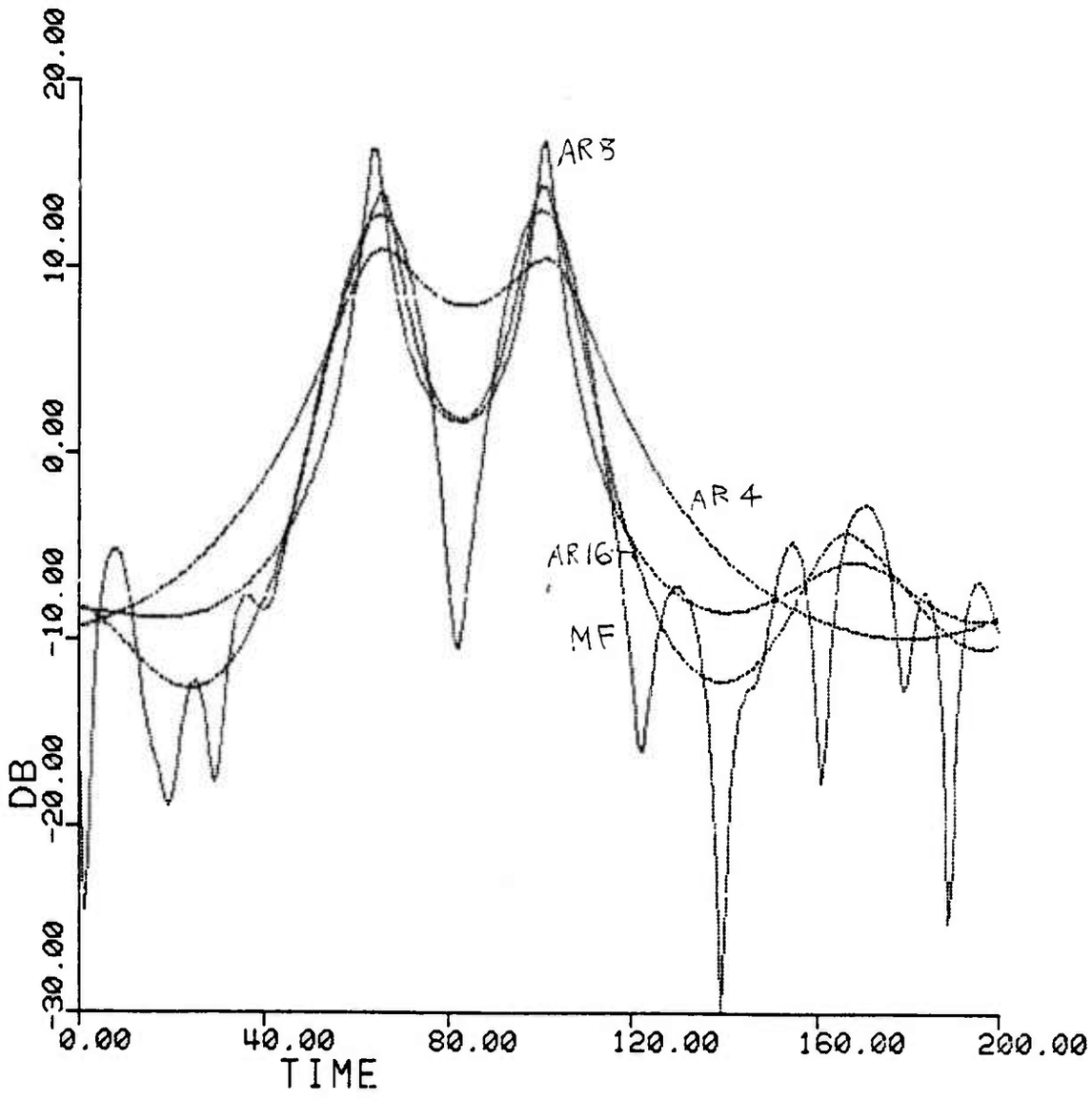


FIG. 2(L)
DELAY ESTIMATION WITH LEVINSON ALGORITHM & MATCHED FILTER. DATA SAME AS IN FIG.25(B)

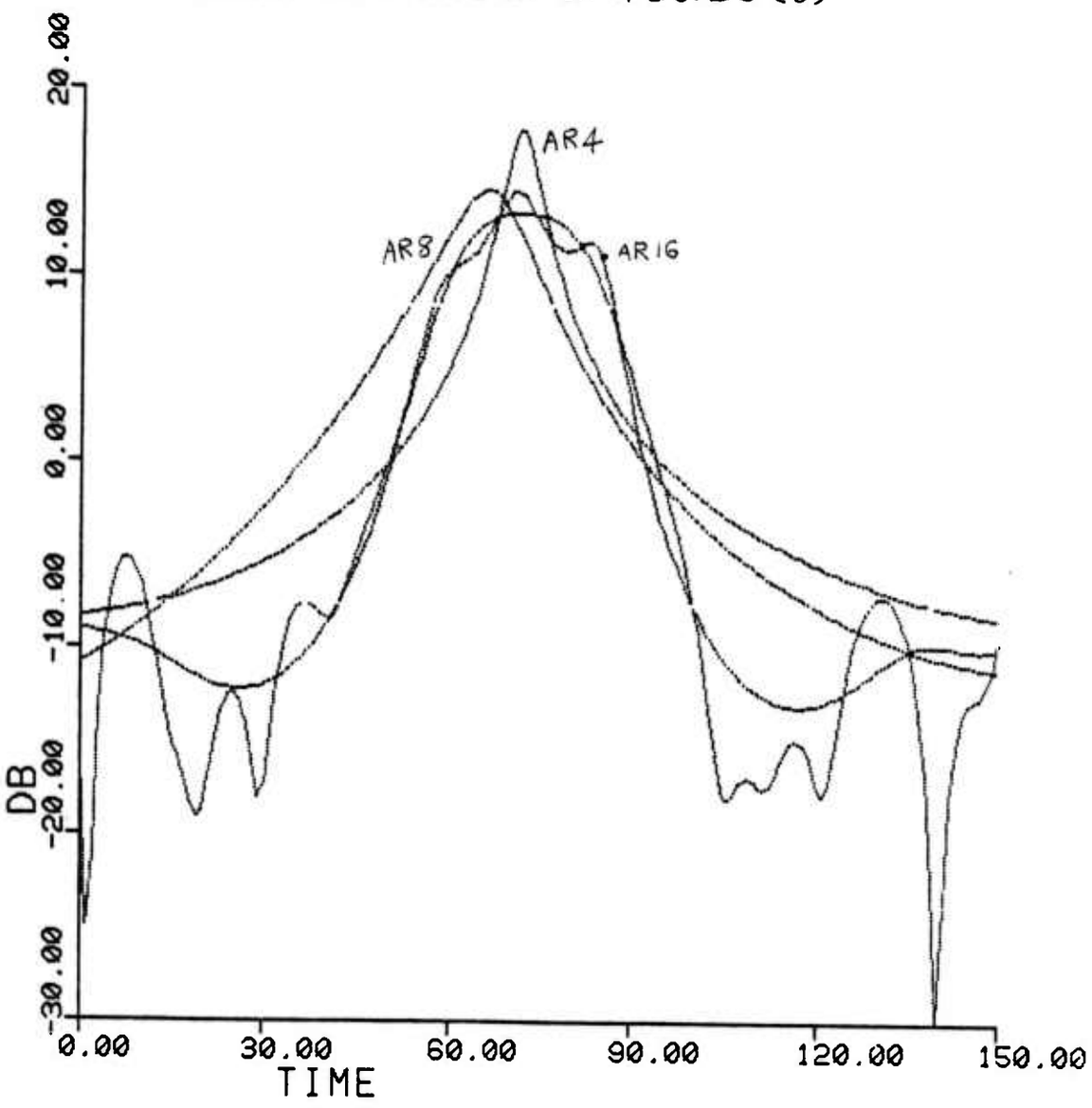


FIG. 3(L)
DELAY ESTIMATION WITH LEVINSON ALGORITHM & MATCHED FILTER. SNR = 30 DB; NUMBER OF SAMPLES = 64; GAUSSIAN ENVELOPE; ONE DELAY AT 72; AR ORDERS 8 & 16.

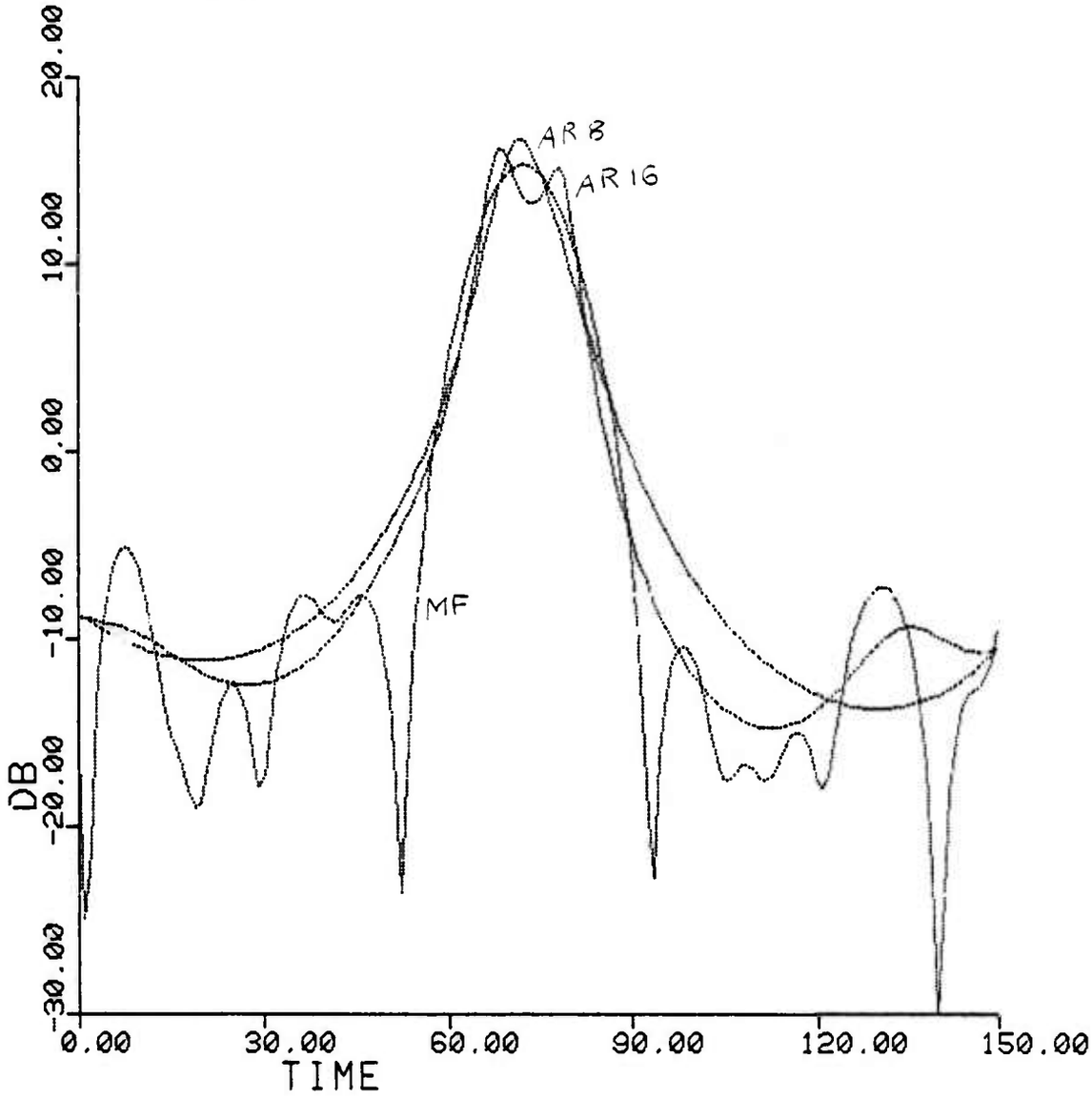


FIG. 4(L)
DELAY ESTIMATION WITH LEVINSON ALGORITHM & MATCHED FILTER. DATA SAME AS IN FIG. 26(B)

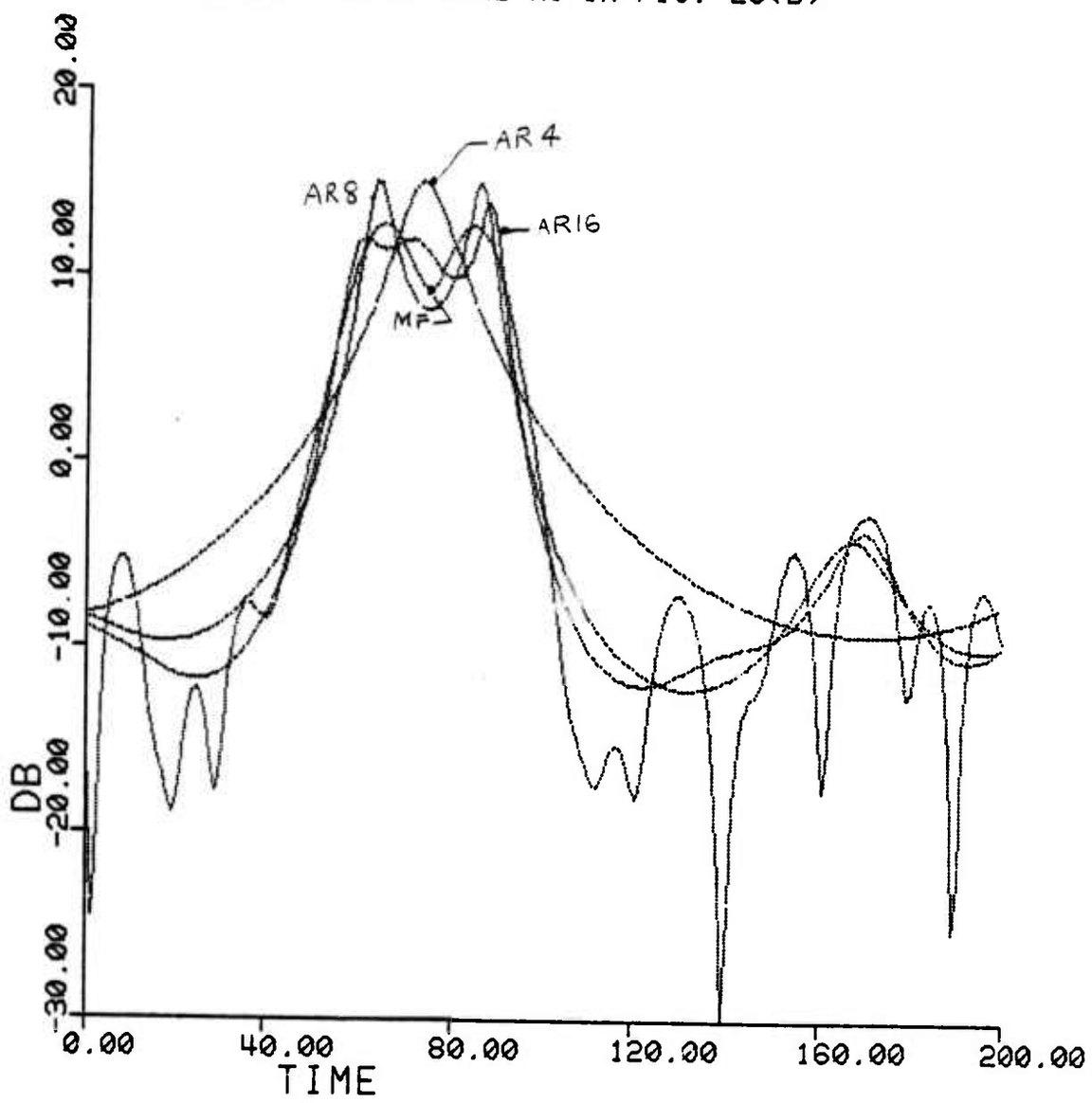


FIG. 5(L)
DELAY ESTIMATION WITH LEVINSON ALGORITHM & MATCHED FILTER; DATA SAME AS IN FIG. 27(B)

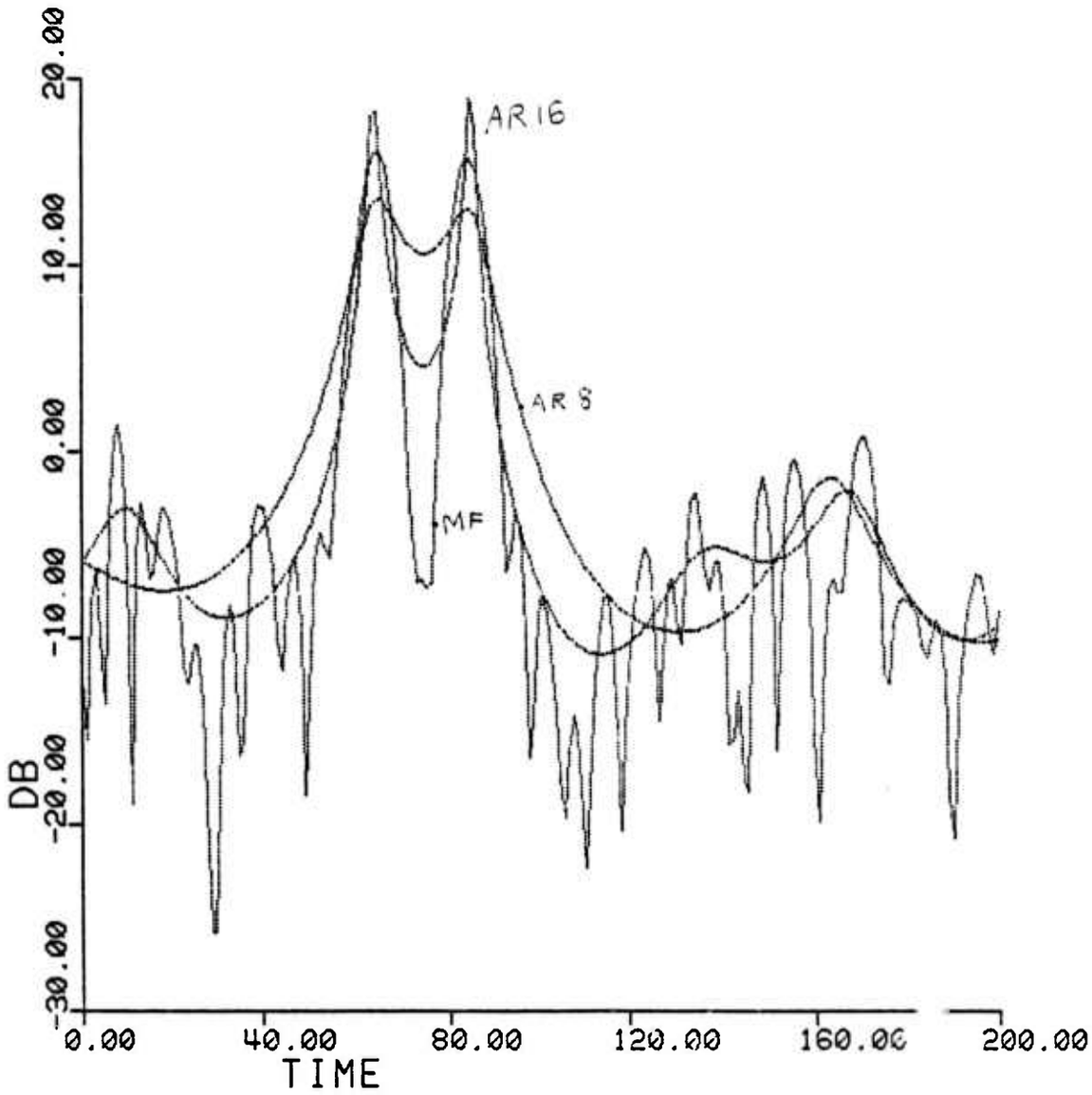
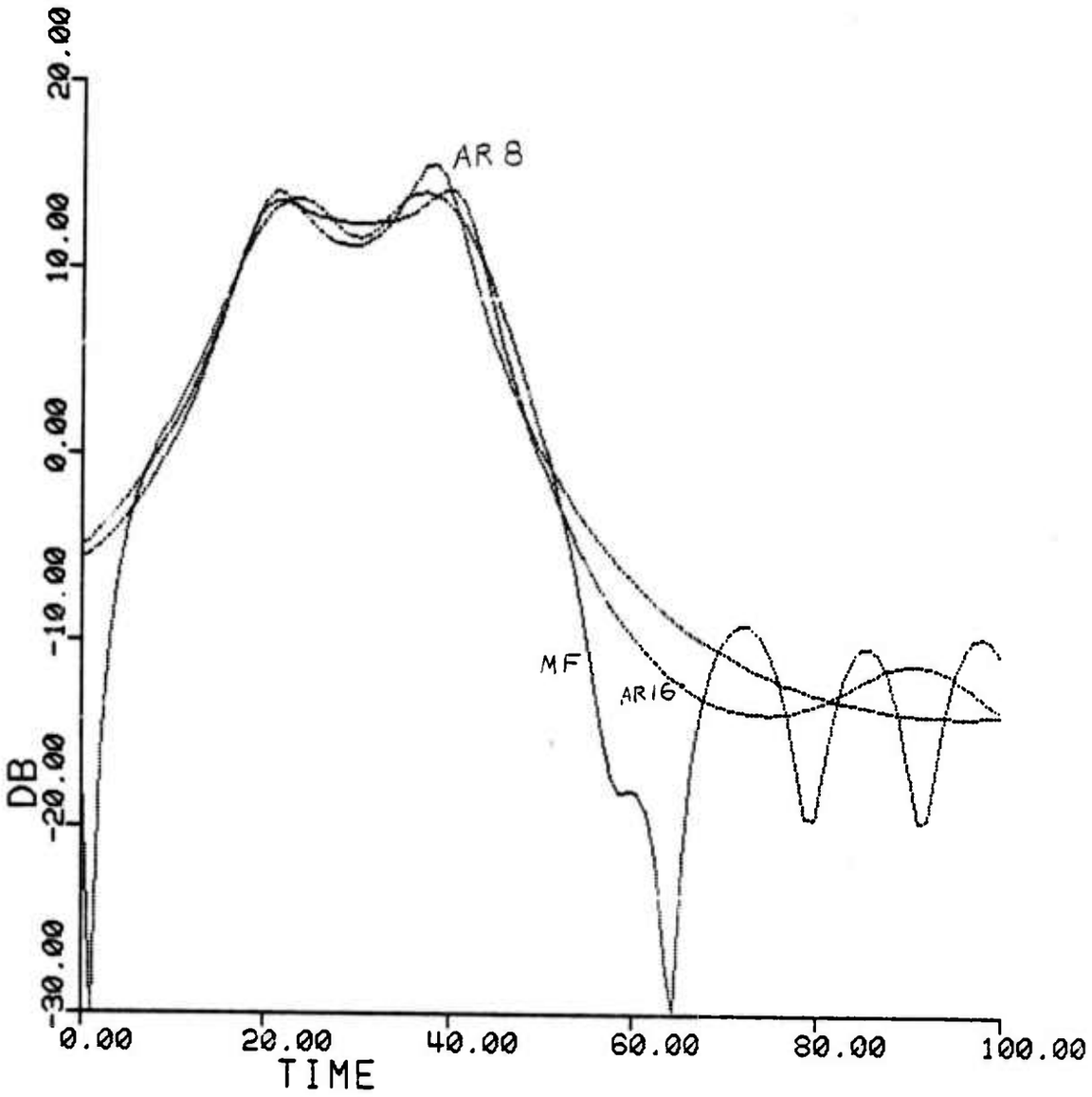


FIG. 6(L)
DELAY ESTIMATION WITH LEVINSON ALGORITHM & MATCHED FILTER; DATA SAME AS IN FIG. 29(B)



DELAY ESTIMATION WITH LEVINSON ALGORITHM & MATCHED FILTER; DATA SAME AS IN FIG. 30(B)

FIG. 7(L)

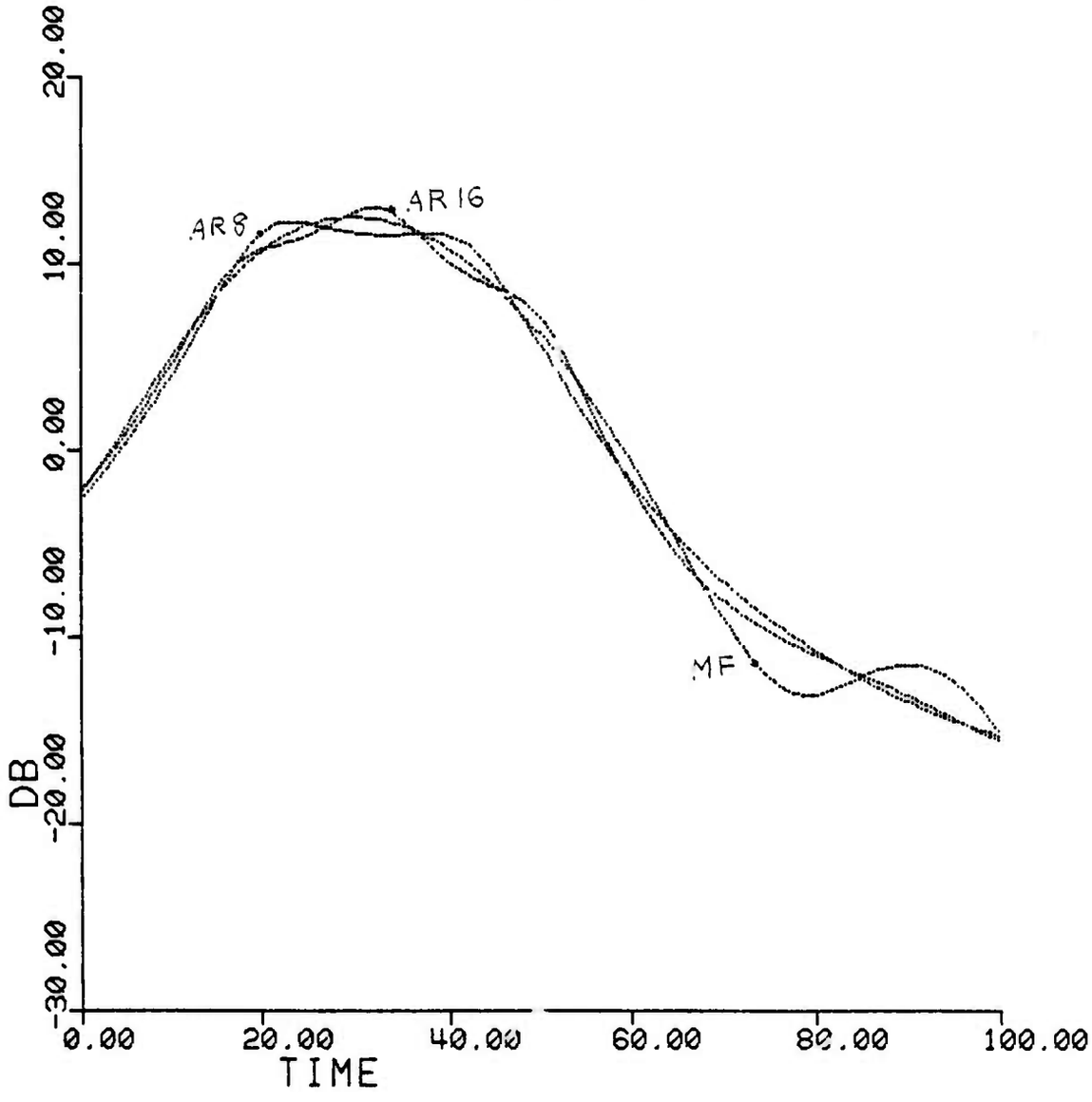
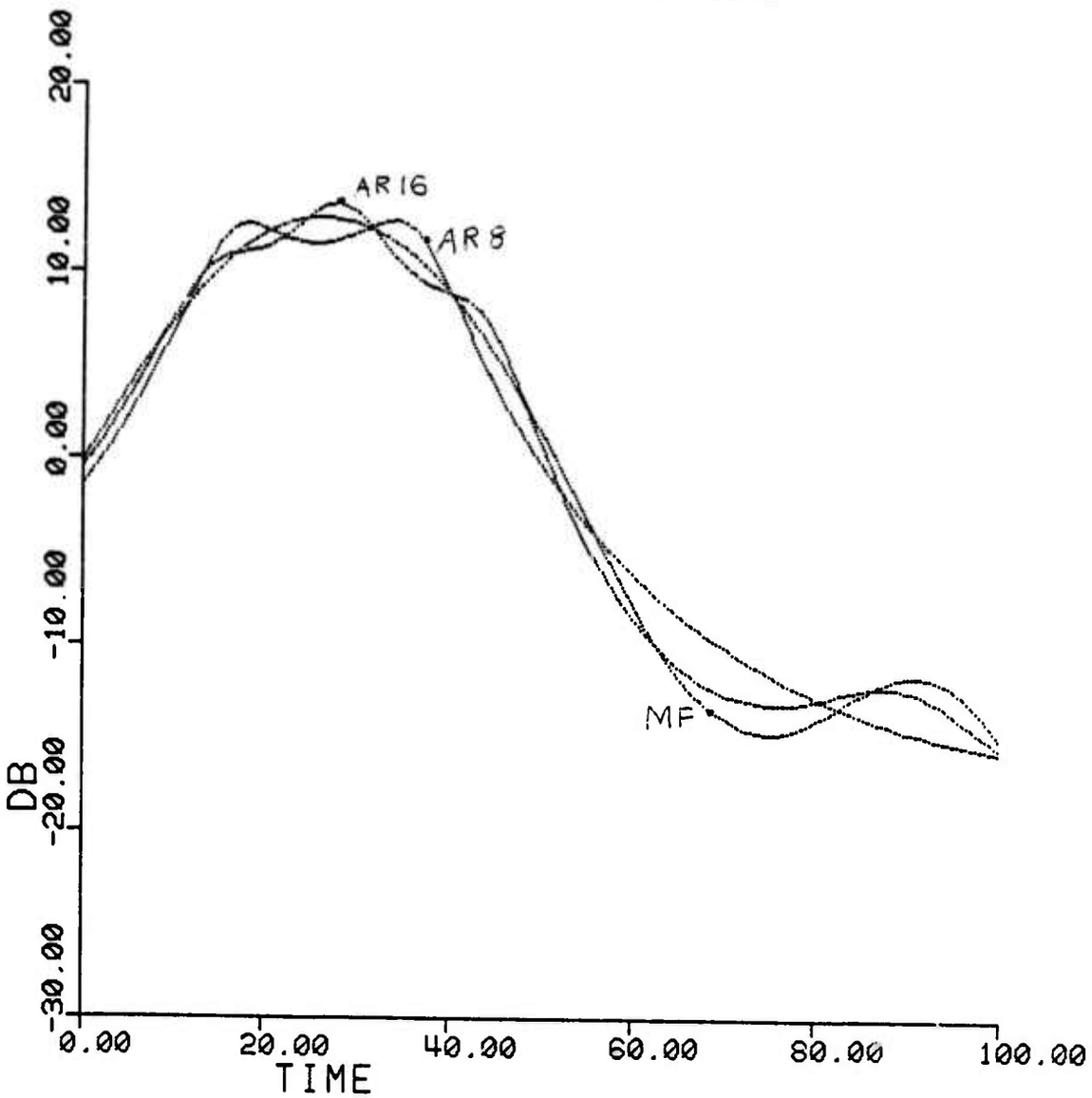


FIG. 8(L)
DELAY ESTIMATION WITH LEVINSON ALGORITHM & MATCHED FILTER; DATA SAME AS IN FIG. 22(B)



Section VIII. Conclusions

While the AR delay estimates appear sharper than the matched filter estimtes when only one echo is present, this property apparently does not translate into the ability to resolve two or more closely spaced echoes reliably. It is easy to display examples where higher resolution is really obtained, but it is also quite possible that a single actual spectral line may have split into two components. Obviously, analogous observations apply to high resolution AR spectral estimation also, by interchanging the roles of the time and frequency domains.

An application of AR modeling to the problem of estimating the instantaneous frequency of a frequency modulated waveform is discussed in Appendix III.

References

1. Akaike, H. "Markovian Representation of Stochastic Processes and Their Application to the Analysis of Autoregressive Moving Average Processes." Ann. Inst. Stat. Math. (Japan), 1974.

2. _____. "Information Theory and an Extension of Maximum Likelihood Principle." Proc. 2nd Int. Symp. on Information Theory, 1971.

3. _____. "Statistical Predictive Identification". Ann. Inst. Stat. Math. (Japan), 1970.

4. _____. "Filtering autoregressive models for prediction." Ann. Inst. Stat. Math. (Japan), 1969.

5. _____. "Power spectrum Estimation Through Autoregressive Model Fitting." Ann. Inst. Stat. Math. (Japan), 1969.

6. Altes, R. A. "Methods of Wideband Signal Design for Radar Sonar Systems." Ph.D. Dissertation, Dept. of Electrical Engineering, Univ. of Rochester, Rochester, NY, 1970.

7. Anderson, T. W. "Statistical Analysis of Time Series." New York: John Wiley, 1971.

8. Berk, K. "Consistent Autoregressive Spectral Estimate." Ann. Stat. Vol. 2, 1974.

9. Blackman, R. B. and J. W. Tukey. "The Measurement of Power Spectra." New York: Dover, 1958.

10. Bogert, B. P., M. J. Healy and J. W. Tukey. "The Frequency Analysis of Time Series for Echoes." in "Time Series Analysis", M. Rosenblatt (Ed.), Wiley, 1963.

11. Bohlin, T. "Comparison of Two Methods of Modeling Stationary EEG Signals." IBM J. Res. Dev. May, 1973.

12. Box, G. E. and G. H. Jenkins, "Time Series Analysis, Forcasting and Control." San Francisco: Holden-Day, 1970.

13. Brillinger, D. R. "Fourier Analysis of Stationary Process." Proc. IEEE, Vol. 62, Dec., 1974.

14. Burg, J. P. "The Relationship between Maximum Entropy Spectra and Maximum Likelihood Spectra." Geophysics, Vol. 37, 1972.

15. _____. "Maximum Entropy Spectral Analysis." Presented at 37th Meeting, Soc. Explor. Geophys., Oklahoma City, Oct. 1967.

16. Capon, J. "High Resolution Frequency Wavenumber Spectrum Analysis." Proc. IEEE, Vol. 57, 1969.

17. Cooley, J. W. and J. W. Tukey. "An Algorithm for the Machine Computation of Complex Fourier Series." Math. Comput. Vol. 19, 1965.

18. Gersch, W. "Estimation of Autoregressive Parameters of a Mixed ARMA Time Series." IEEE Trans. Auto. Control. Vol. AC-15, 1970.

19. Graupe, D. "Estimation of Upper Bounds of Errors in Identifying Autoregressive Models." Int. J. Syst. Sci. Vol. 6, 1975.

20. _____. and J. Perl. "Stochastic Approximation Algorithms for Identifying ARMA Processes." Int. J. Syst. Sci. Vol. 5, 1974.

21. Griffiths, L. J. "Rapid Measurement of Digital Instantaneous Frequency." IEEE Trans. Vol. ASSP. 23, 1975.

22. Gupta, S. C. "Phase Locked Loops." Proc. IEEE, Feb. 1975.

23. Ionnides, G. A. "Application of Multivariate Autoregressive Spectral Estimation to ULF Waves." Rad. Sci., Vol. 10, Dec. 1975.

24. Jones, R. H. "Identification and Autoregressive Spectral Estimation." IEEE Trans. Auto. Control, Vol. AC-19, Dec. 1974.

25. Kailath, T. (Ed). Special Issue on System Identification and Time Series Analysis. IEEE Trans. Vol. AC-19, Dec. 1974.

26. Kaveh, M. and G. R. Cooper "Empirical Investigation of the AR Spectral Estimator." IEEE Trans. Vol. IT.

27. Koopmans, L. H. "Spectral Analysis of Time Series." New York, Academic Press, 1974.

28. Lacoss, R. T. "Data Adaptive Spectral Analysis Methods." Geophysics, Vol. 36, 1971.

29. Levinson, N. "The Wiener RMS Error Criterion in Filter Design and Prediction." J. Math. Phys. Vol. 25, 1947.

30. Makhoul, J. "Spectral Analysis of Speech by Linear Prediction." IEEE Trans. Vol. AU-21, June 1973.

31. _____. "Linear Prediction - A Tutorial Review." Proc. IEEE, Vol. 63, April 1975.

32. Parzen, E. "On Consistent Estimates of the Spectrum of Stationary Time Series." Ann. Math. Stat. Vol. 28, 1975.

33. _____. "Multiple Time Series Modeling" in "Multivariate Analysis." P. R. Krishnaiah (Ed.), New York, Academic Press, 1969.

34. _____. "Some Recent Advances in Time Series Modeling." IEEE Trans. Vol. AC-19, Dec. 1974.

35. _____. "An Approach to Time Series Analysis." Ann. Math. Stat. Vol. 32, 1961.

36. Peacock, K. L. and S. Treitel. "Predictive Deconvolution - Theory and Practice." Geophys. Vol. 34, 1969.

37. Pisarenko, V. E. "On the Estimation of Spectra by Means of Non-linear Functions of the Covariance Matrix." Geophys. J. Royal Astron. Soc. Vol. 28, 1972.

38. Proakis, J. "Equalization for Intersymbol Interference." (in) Advances in Communications Theory, A. V. Balakrishnan (Ed.), New York: Academic Press, 1975.

39. Rabiner, L. R. And B. Gold. "Theory and Applications of Digital Signal Processing." Englewood Cliffs, NJ: Prentice-Hall, 1974.

40. Radoski, H. R., P. F. Fougere and E. J. Zawalick. "A comparison of Power Spectral Estimates and Applications of Maximum Entropy Method." J. Geophy. Res. Vol. 80, 1975.

41. Rao, C. R. "Linear Statistical Inference and its Applications." New York: John Wiley, 1973.

42. Tong, H. "Autoregressive Model Fitting with Noisy Data by Akaike's Information Criterion." IEEE Trans. Vol. IT-21, 1975.

43. Tretter, S. A. and K. Steiglitz. "Power Spectrum Identification in terms of Rational Models." IEEE Trans. Vol. AC-12, 1967.

44. Ulrych, T. "Maximum Entropy Power Spectrum of Truncated Sinusoids." J. Geophys. Res. Vol. 77, 1972.

45. Viterbi, A. J. "Principles of Coherent Communication." New York: Academic Press, 1965.

46. Ulrych, T. and T. N. Bishop. "Maximum Entropy Spectral Estimation and AR Decomposition." Rev. Geophys. Space - Phy. 1975.

47. Van Den Bos, A. "Alternative Interpretation of Maximum

Entropy Spectral Analysis." IEEE Trans. Vol. IT-17, July 1971.

48. Van Trees, H. L. "Detection, Estimation and Modulation Theory, Vol. I." New York: Wiley, 1968.

49. Wang, R. J. and S. Treitel. "The Determination of Digital Wiener Filters by means of Gradient Methods." Geophys. Vol. 38, 1973.

50. Wang, R. J. "Optimum Window Length for the Measurement of Time Varying Power Spectra." J. Acoust. Soc. Am., Vol. 52, 1971.

51. Widrow, B. "Adaptive Filters - I: Fundamentals." Stanford Electronics Lab., Rept. SEL-66-126. Stanford, Calif. 1966.

52. Widrow, B. et. al. "Adaptive Noise Cancellation - Principles and Applications." Proc. IEEE, Vol. 63, 1975.

53. Private Communication from E. Titlebaum, University of Rochester, Rochester, N.Y.

54. Oppenheim, A. V. and R. Schafer. "Digital Signal Processing." Prentice-Hall, 1975.

55. Abramowitz, M. and I. Stegun. "Handbook of Mathematical Functions", New York: Dover, 1965.

Appendix - A.1

The Levinson Algorithm [29].

The Levinson algorithm computes the solution of the equations

$$\underline{\underline{A}}\ \underline{C} = \underline{v}, \qquad \text{(A.1.1)}$$

where $\underline{\underline{A}}$ is a Toeplitz Hermitian (LxL) matrix, in a number of steps proportional to L^2.

$$\text{Let } \underline{\underline{A}} = \begin{Bmatrix} 1 & \gamma_1 & \gamma_2 & \cdots & \gamma_{L-1} \\ \gamma_1^* & 1 & \gamma_1 & \cdots & \gamma_{L-2} \\ \vdots & & & & \\ & & & 1 & \gamma_1 \\ \gamma_{L-1}^* & & & \gamma_1^* & 1 \end{Bmatrix} ;$$

$$\underline{\gamma}_i = (\gamma_1, \gamma_2, \ldots \gamma_i)^T$$

$$\underline{\bar{\gamma}}_i = (\gamma_i, \gamma_{i-1}, \ldots \gamma_2, \gamma_1)^T$$

$$\underline{v}_i = (v_1, v_2, \ldots v_i)^T$$

$$\underline{\underline{A}}_i = \begin{matrix} 1 & \gamma_1 & \gamma_2 & \cdots & \gamma_{i-1} \\ \gamma_1^* & 1 & \gamma_1 & \cdots & \gamma_{i-2} \\ \vdots & & & & \\ \gamma_{i-1}^* & & & \gamma_1^* & 1 \end{matrix}$$

and

$$\underline{C}_i = \underline{\underline{A}}_i^{-1}\ v_i.$$

Then using matrix inversion lemmas, the following algorithm can be derived:

$$\underline{C}_1 = v_i$$

$$\gamma_1 = 1 - |\gamma_1|^2$$

$$g_1 = \gamma_1$$

For i = 1,2, . . . L-1 do:

$$\mu_i = (\overline{\underline{\gamma}}_i^{*T}\, \underline{C}_i - v_{i+1})/\lambda_i$$

$$\eta_i = (\underline{\gamma}_i^T\, \underline{g}_i - \gamma_{i+1})/\lambda_i$$

$$\underline{D}_{i+1} = \begin{matrix} \underline{C}_i + \underline{g}_i\, \mu_i \\ - - - - \\ -\mu_i \end{matrix}$$

$$\lambda_{i+1} = \lambda_i(1-|\eta_i|^2)$$

$$\underline{g}_{i+1} = \begin{matrix} -\eta_i \\ - - - - \\ \underline{g}_i + \eta_i\, \overline{\underline{g}}_i^* \end{matrix}$$

$$\overline{\underline{g}}_{i+1} = \underline{g}_{i+1} \text{ read in backward order}$$

$$= (g_{i+1,\, i+1}\ \ g_{i,\, i+1} \cdot \cdot \cdot g_{1,i+1})$$

C_L = solution of the equation $\underline{A}\ \underline{c} = \underline{v}$.

Appendix - A II

The complex data $\{x(i)\}_{i=1}^{N}$ is modeled by the autoregressive representation

$$x(k) = \sum_{i=1}^{L} C_i \, x(k-j) + n(k)$$

The Burg algorithm generates estimates of the PEF coefficients without explicitly calculating the data ACF beforehand. The procedure is analogous to the Levinson algorithm.

Consider the equations, with $1 \leq k \leq L$,

$$\begin{Bmatrix} \hat{\gamma}_0 & \hat{\gamma}_1 & \cdots & \hat{\gamma}_k & 1 \\ \hat{\gamma}_1^* & \hat{\gamma}_0 & \cdots & & -C_{k,1} \\ \cdots & \cdots & \cdots & \hat{\gamma}_0 \; \hat{\gamma}_1 & -C_{k,2} \\ \hat{\gamma}_L^* & \hat{\gamma}_{L-1}^* & \cdots & \hat{\gamma}_1^* \; \hat{\gamma}_0 & -C_{k,k} \end{Bmatrix} = \begin{Bmatrix} P_{k+1} \\ 0 \\ \cdot \\ \cdot \\ 0 \end{Bmatrix} \qquad \text{(A.2.1)}$$

For k=L, (A.2.1) are the normal equations for the estimation of the PEF coefficients. The ACF estimates $\hat{\gamma}(i)$, i=0, . . . L are presently unknown.

Let $P_1 = \frac{1}{N} \sum_{i=1}^{N} |x(i)|^2$

Partition vector $(1, -C_{k,1}, -C_{k,2}, \ldots -C_{k,k})^T$ and $(P_{k+1}, 0, 0, \ldots)$ into

(equation on next pg.)

$$\begin{Bmatrix} 1 \\ -C_{k,1} \\ \cdot \\ \cdot \\ \cdot \\ -C_{k,k} \end{Bmatrix} = \begin{Bmatrix} 1 \\ -C_{k-1,1} \\ -C_{k-1,2} \\ \cdot \\ \cdot \\ -C_{k-1,k-1} \\ 0 \end{Bmatrix} - C_{k,k} \begin{Bmatrix} 0 \\ -C^*_{k-1,k-1} \\ -C^*_{k-1,k-2} \\ \cdot \\ 1 \end{Bmatrix} \qquad \text{(A.2.2)}$$

and

$$\begin{Bmatrix} P_{k+1} \\ 0 \\ \cdot \\ \cdot \\ 0 \end{Bmatrix} = \begin{Bmatrix} P_k \\ \cdot \\ 0 \\ \cdot \\ \Delta_k \end{Bmatrix} - C_{k,k} \begin{Bmatrix} \Delta^*_k \\ \cdot \\ 0 \\ \cdot \\ P_k \end{Bmatrix} \qquad \text{(A.2.3)}$$

from these,

$$\Delta_k = C_{k,k}\, P_k \qquad \text{(A.2.4)}$$

$$P_{k+1} = P_k(1-|C_{k,k}|^2) \qquad \text{(A.2.5)}$$

$$\hat{\gamma}(k) = \Delta^*_k + \sum_{i=1}^{k-1} \hat{\gamma}(i)\, C^*_{k-1,k-i} \qquad \text{(A.2.6)}$$

$$C_{k,i} = C_{k-1,\,i} - C_{k,k}\, C^*_{k-1,k-i} \text{ for } 1 \le i \le k-1 \qquad \text{(A.2.7)}$$

If $\hat{\gamma}(k)$ were precomputed, $C_{k,k}$ can be determined from (A.2.4) and (A.2.6). This gives one version of Levinson's algorithm

for the computation of AR coefficients, as can be verified by comparison with the algorithm given in Appendix - I.

In the Burg algorithm, $C_{k,k}$ is determined by minimizing

$$P_{\ell,k} = \sum_{i=1}^{N-k} |x_{(i)} - \sum_{\ell=1}^{k} C^*_{k,\ell} x_{(i+\ell)}|^2 + |x_{(i+k)} - \sum_{\ell=1}^{k} x_{(i+k+\ell)} C_{k,}|^2 \qquad \text{(A.2.8)}$$

This is equivalent to running the k-th order PEF over the data in forward and backward directions and computing the sum of squares of residuals.

Define

$$a_{i,k} = x(i) - \sum_{\ell=1}^{k-1} C^*_{k-1,\ell}\, x(i+\ell)$$

$$b_{i,k} = x(i+k) - \sum_{\ell=1}^{k-1} C_{k-1,k-\ell}\, x(i+\ell) \qquad \text{(A.2.9)}$$

Then using (A.2.7) and (A.2.9), (A.2.8) can be written as

$$P_{\ell,k} = \sum_{i=1}^{N-k} |a_{i,k} - C^*_{k,k}\, b_{i,k}|^2 + |b_{i,k} - C_{k,k}\, a_{i,k}|^2$$

The value of $C_{k,k}$ that minimizes this expression is

$$C_{k,k} = \frac{\sum_{i=1}^{N-k} a^*_{i,k}\, b_{i,k}}{\sum_{i=1}^{N-k} |a_{i,k}|^2 + |b_{i,k}|^2} \qquad \text{(A.2.10)}$$

Equations (A.2.5), (A.2.7) and (A.2.10), together with the initial value $P_1 = \frac{1}{N} \sum_{i=1}^{N} |x(i)|^2$, give the Burg algorithm. The iterations are performed for k=1, 2, . . . L and it is assumed that any variable with a zero subscript has the value of zero.

Appendix - III

Instantaneous Frequency estimation of wideband frequency modulated waveforms using AR modeling.

If the frequency of a noisy sinusoidal input is time invariant and an adequate number of samples is available, the usual Fourier Transform methods offer a combination of near optimality in a decision theoretic sense, good resolution and efficient computational algorithms. If the frequency varies in a narrow band around a known carrier frequency, phase locked loop techniques should be considered [22,45]. AR frequency estimation techniques appear to be effective when the input frequency varies in a wideband over the Nyquist frequency range. If the input frequency varies slowly with time, the effective time interval over which the frequency is nearly constant is limited. In such situations, the ability of the AR spectral estimator to produce good frequency estimates with short data lengths is advantageous.

Griffiths [21] has noted that the ability of the LMS algorithm to track slow time variations in regression coefficient estimates can be put to use to estimate short term spectra. However, this scheme does not produce estimates of the instantaneous frequencies themselves. The task of recovering the instantaneous frequency from the PEF coefficients still remains. This is usually done by computing the AR PSD using a DFT or its equivalent and selecting the frequency at which it attains a maximum. Even though the AR LMS algorithm is, in principle, capable of updating the input frequency estimate at every sampling instant, in practice it is feasible to do so only at widely spaced time instants.

An alternative approach is to collect a small block of samples at a periodic rate determined by the rate of variation of the input frequency, and, assuming that the frequency does not change within each block, determine the frequency at the block sampling instants using Burg's algorithm and a DFT. That is, input samples $\{x(kT + n)\}$, $k = 0, 1, 2, \ldots$ and $n = (1, 2, \ldots N)$ with $T>>N$, are analyzed with Burg's Algorithm of order L, $L<<N$, and the resulting frequency estimate is held to be the input frequency at time kT. By suitable interpolation, the value of f(t) for other values of time t can be determined.

Two simulations were run to compare the computational times required for the Burg and LMS approaches. In the first example, a frequency modulated waveform was generated by the formula

$$x(t) = \cos\left(\frac{2\pi t}{16}\left(1+\frac{t}{10000}\right)\right) \text{ for } 0 \leq t \leq 5000$$

$$= \cos\left(\frac{2\pi t}{16}\left(3-\frac{t}{10000}\right) - \frac{10000\pi}{16}\right), \quad 5000 \leq t \leq 10032.$$

White noise of variance 0.01 (corresponding to a carrier to noise ratio of 17db) was added to x(t), and the sampling rate was 1 Hz.

The instantaneous freuqency of x(t) is

$$f(t) = \frac{1}{16}(1+t/5000), \qquad 0 \leq t \leq 5000$$

$$= \frac{1}{16}(3 - t/5000), \qquad 5001 \leq t \leq 10032.$$

The input samples were processed with a 4th order AR-LMS algorithm with $\alpha = 0.04$ (eq. (3.3)). After every 500 iterations, the instantaneous frequency was updated from the 256 point DFT of the PEF coefficients, as discussed above. The times required for the computations on an Interdata 7/32 machine were 13±1 seconds for the LMS algorithm and 9 seconds for the DFT calculations.

The same data were processed with a 4th order Burg's algorithm after sampling blocks of 32 points each, every 200 seconds. (T=200, N=32, L=4). The total times required on the same machine were: 4 ± 1 sec. for the Burg algorithm and 9 seconds for the DFT. The actual instantaneous frequencies of the input, as well as the estimates derived from the LMS and Burg algorithms, are plotted in Fig. 2(F).

In the second example, the input waveform is described by

$$x(t) = \cos\,(250 \sin\,(2\pi t/2000)) + n(t), \quad 0 \leq t \leq 10032$$

with variance of n(t) being 0.01 and the sampling interval, 1 sec. The instantaneous frequency function in this case is

$$f(t) = \frac{1}{8} \cos\,(2\pi t/2000).$$

This data was processed with the AR-LMS algorihtm (L-4, = 0.04) and Burg's algorithm (L-4, N=32, T=200). In either case, updating of the instantaneous frequency estimates was performed every 200 seconds. The LMS algorithm required 13 ± 1 seconds for the AR coefficient computations and the

Burg algorithm took 5 $\pm$ 1 seconds on the same Interdata machine. The time required for the DFT and peak selection procedures was 22 $\pm$ 1 sec. in either case. The results are plotted in Fig. 4(F).

Discussion

Both methods give reasonable good estimates of the input frequency. The LMS estimates have a somewhat lower error than the Burg estimates, though the Burg algorithm is more than twice as fast as the LMS algorithm for these two examples.

We do not claim that the Burg algorithm approach is necessarily superior to the LMS approach. At lower SNRs it is likely that the LMS algorithm performs much better. If the input frequency estimates have to be updated more frequently, the LMS algorithm will become computationally superior. On the other hand, the LMS algorithm could become unstable for bad choices of α [21], while the Burg algorithm is guaranteed to be stable. At high SNR and with slow variations in the input frequency, the Burg algorithm is an attractive alternative to the LMS algorithm.

FIG. 2(F)
INSTANTANEOUS FREQUENCY ESTIMATION WITH BURG & LMS ALGORITHMS. (SEE TEXT FOR DATA).
SOLID LINE : ACTUAL VALUES; ASTERISKS : BURG; SQUARES : LMS.

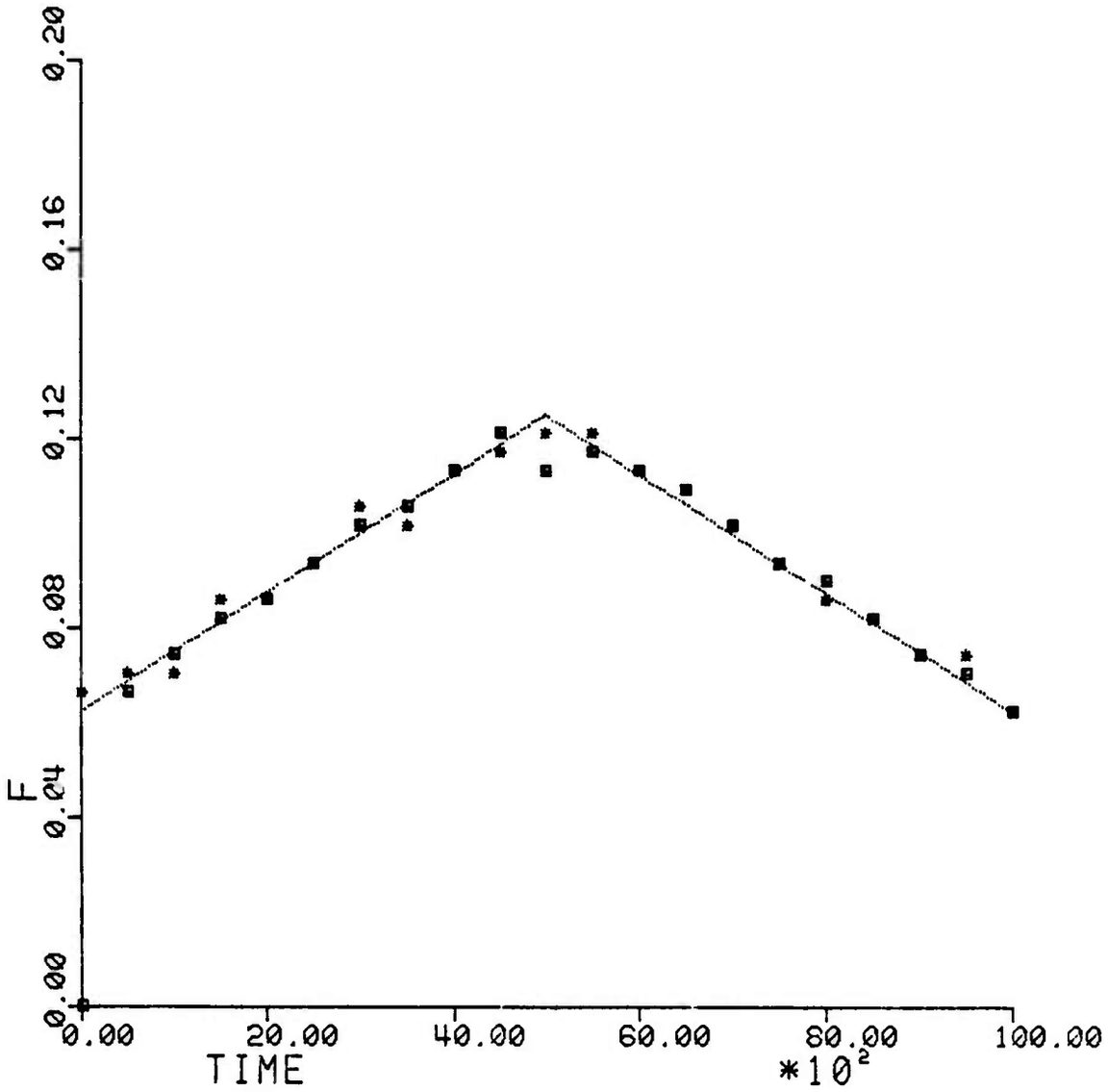
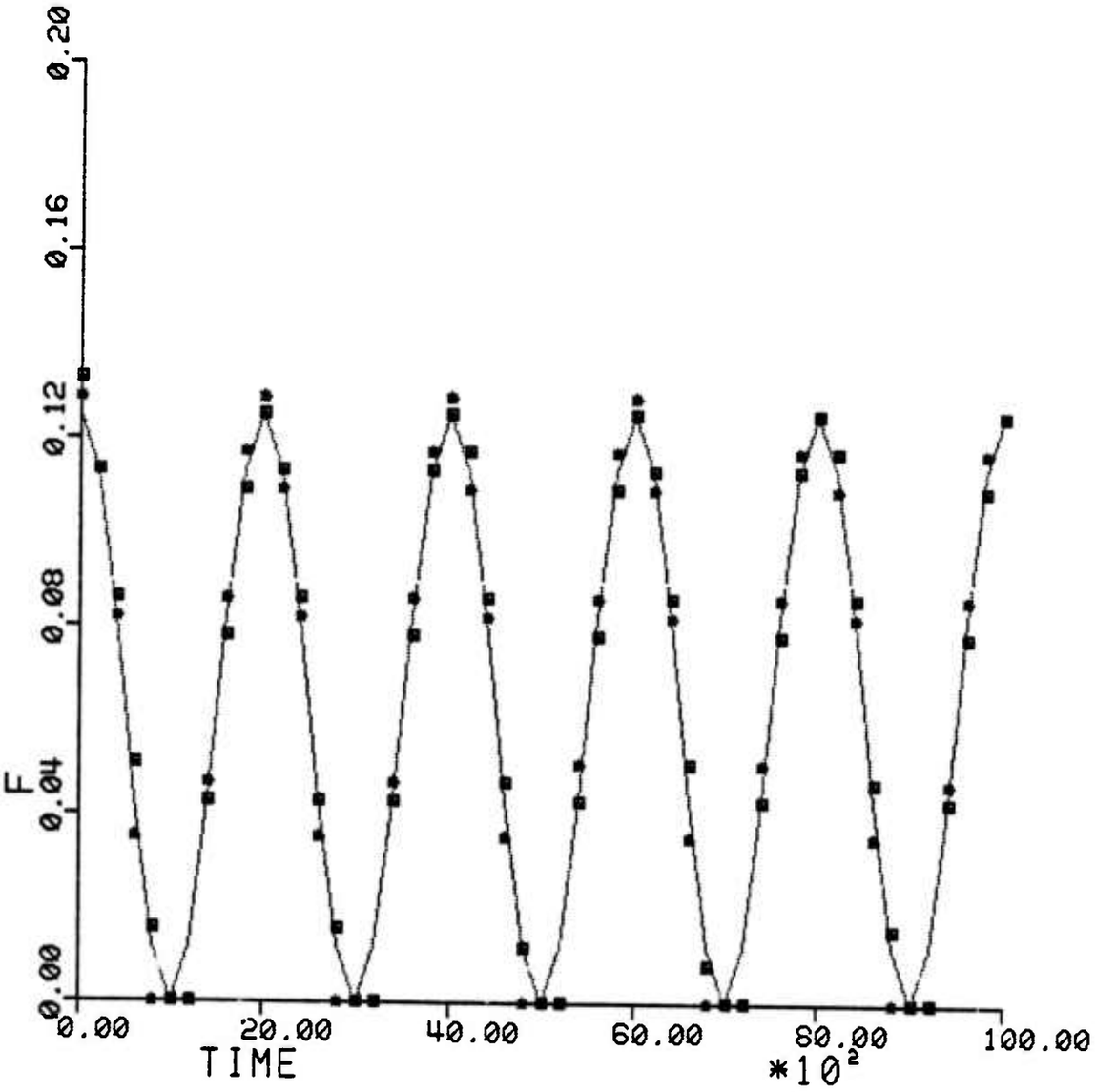


FIG. 4(F)
INSTANTANEOUS FREQUENCY ESTIMATION WITH BURG & LMS ALGORITHMS. (SEE TEXT FOR DATA.)
SOLID LINE : ACTUAL VALUES; ASTERISKS : BURG; SQUARES : LMS.



ADAPTIVE ESTIMATION OF TIME VARYING PARAMETERS IN UNDERSEA ACOUSTIC COMMUNICATION CHANNEL MODELS

by

S. NARAYANA MURTHY

GRADUATE ASSISTANT

Dept. of Electrical Engineering

University of Rochester

Rochester, N.Y. 14627

Notation

The following symbols, conventions and abbreviations will be used through this chapter.

ALMS : Approximate Least Mean Square algorithm.

$\underline{c}_n$: Unknown time varying parameters to be estimated.

$E\{\ \}$: Expectation operator.

ep_n : Prediction error at time n = $x_n - \underline{s}_n^T \hat{\underline{c}}_{n-1}$

EWLS : Exponentially weighted Least Squares Algorithm

f : frequency variable

$\underline{I}$: Identity matrix

j : $\sqrt{-1}$

L : dimension of $\underline{c}_n$ (numbers of unknown parameters)

LMS : Widrow's Least Mean Square Algorithm [51].

$\underline{M}$: diagonalizing matrix of $\underline{R}_{ss}$; $M^{-1} = M^{*T}$

i.e. $\underline{M}\ \underline{\Lambda}\ \underline{M}^{-1} = \underline{R}_{ss}$

$\underline{N}_u$: Time varying component of $\underline{P}_n$.

$\underline{P}_n^{-1}$: $\sum_{k=1}^{w} (\underline{I}-\underline{Q})\ \underline{Q}^{n-k}\ \underline{s}_k^*\ \underline{s}_k^T$ = estimated signal autocorrelation matrix.

q : exponential weight.

$\underline{Q}$: weighting matrix; all the eigenvalues of $\underline{Q}$ are between 0 and 1.

$\underline{R}_{ss}$: input autocorrelation matrix.
$= \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \underline{s}_k^*\ \underline{s}_k^T$. $\underline{R}_{ss}$ is assumed to be positive definite and Hermitian.

$\underline{R}_{sx}$: cross correlation between input vector and observations. $= \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \underline{s}_k^*\ x_k$

$\{\underline{s}_k\}$: set of known input signal vectors.

t : continuous time variable.

$\{x_k\}$: set of scalar observations.

$\{w_k\}$: set of additive noise samples; assumed to white with variance σ^2.

i.e. $E\{w_k\ w_l^*\} = \delta_{k,l}\ \sigma^2$.

α : gain constant in ALMS algorithm.

$\underline{\underline{\Lambda}}$: matrix of eignevalues of $\underline{\underline{R}}_{ss}$.

σ^2 : additive noise variance

$\underline{P}$: doppler scale vector.

Double underlining denotes a matrix.

Single underlining denotes a vector.

$[\underline{\underline{\lambda}}_i]$ denotes a diagonal matrix whose $(i,i)^{th}$ element is λ_i.

Superscripts

$*$: denotes complex conjugate

T : denotes transpose

$\hat{}$: estimate

$\sim$: Fourier Transform

-1 : Matrix inverse

e.g. $\underline{\tilde{\hat{c}}}(f)^{*T}$ would mean the complex conjugate transpose of the Fourier transform of the estimate of $\underline{c}(f)$.

$\underline{\underline{Q}}^k$: k^{th} power of $\underline{\underline{Q}}$.

$\{\underline{\underline{Q}}^k\}$: a matrix valued low pass filter whose impulse response is $(\underline{\underline{I}}, \underline{\underline{Q}}, \underline{\underline{Q}}^2, \ldots)$

Subscripts and parentheses

subscript n : value of indicated function at time n

e.g. $\underline{s}_n$ means the input signal vector at time n.

parentheses i,1 : (i,1) component of the indicated matrix

e.g. $s_n(i)$ = the i^{th} component of $\underline{s}_n$.

parentheses t : value of the continuous function at time t.

e.g. s(1)(P(1)·t) means value of the first component of $\underline{s}(t)$ at time P(1)·t.

$|\cdot|^2$: Euclidean norm e.g. $|\underline{s}_n|^2 = \sum_{k=1}^{L} |s(k)|^2$.

Differentiation with respect to vector:

$$\frac{\partial f}{\partial \underline{c}} = \left(\frac{\partial f}{\partial c(1)}, \frac{\partial f}{\partial c(2)}, \frac{\partial f}{\partial c(3)}, \cdots \frac{\partial f}{\partial c(L)}\right)^T$$

I. Introduction

We resume the study of identifying underwater acoustic channel models of the form

$$x(t) = \sum_{i=1}^{L} c(i)\ (t)\ s(P(i)\ (t-\tau(i))) + {}^{w(t)} \qquad (1.1)$$

Knowing $x(t)$, $s(t)$ and $\{\tau_i\}$, the problem is to identify the time varying coefficients $c(i)\ (t)$. The doppler scales $\{P(i)\}$ are unknown, and it is desirable that the estimates of $c(i)\ (t)$ are insensitive to the values of $\{P(i)\}$.

As in the previous chapter, we assume that all waveforms are sampled with a period of 1 sec., and that the samples will be digitally processed. All aliasing errors are assumed to be negligible. With these assumptions, we get the discrete model

$$x_n = \underline{c}_n^{\,T}\ \underline{v}_n + w_n, \qquad \text{where} \qquad (1.2)$$

$$\underline{c}_n = (c(1)(t),\ c(2)(t)\ \ldots\ c(L)(t)\big|_{t=n})^T$$

$$\underline{v}_n = \{s(P(1)(t-\ (1))),\ \ldots\ s(P(L)(t-\tau(L)))\}\Big|_{t=n}^{T}$$

$\{w_n\}$ is assumed to be a white noise sequence with variance σ^2 i.e. $E\{w_n w_k^*\} = \delta_{n,k}\ \sigma^2$.

At first, we shall set $\underline{v}_k = \underline{s}_k$ i.e. no doppler shift will be assumed to be present. This restriction will be later removed.

Technically, the assumption that $\{\tau(i)\}$ are known is not necessary. The channel can be modeled as a tapped delay line, the number of taps being equal to twice the time bandwidth product, 2TB, of the channel [37]. Then equation (1.1) can be replaced by

$$x(t) = \sum_{i=1}^{2TB} c(i)\ (t)\ s(P(i)(t-i))) + w(t) \qquad (1.1a)$$

If approximate a priori estaimtes of $\{\tau_i\}$ are available, the number of unknown parameters can be considerably reduced.

The Widrow-Hoff LMS algorithm has been used successfully in many instances to estimate time varying regression parameters [40,52,53, 54]. In most of these cases, the input signals are assumed to be uncorrelated pseudorandom noise. Indeed, it has proved difficult to analyze the operation of the LMS algorithm with other kinds of input signals. Daniell [9] has shown that with stochastic input signals with strong mixing property (i.e. whose autocorrelation function tend to zero at increasing values of lag in a prescribed manner), the variance of the parameter estimates using the LMS algorithm remains bounded for slow adaption rates. Glover [22] has studied the operation of the LMS algorithm with sinusoidal inputs.

Widrow and his coworkers have done extensive simulation and experimental studies with the LMS algorithm and suggested some useful heuristic analytic models to describe its properties.

Much of the difficulty in analyzing the performance of the LMS algorithm stems from viewing it as a variant of the Stochastic

approximation (S.A.) algorithms [46]. This point of view, while undoubtedly valid, is severely restrictive. Most of the work on S.A. has focused on formal proofs of asymptotic convergence of certain iterative schemes for solving nonlinear regression equations with conditionally independent observations. These proofs are not generalized easily by relaxing some of the assumptions made there in. As far as linear channel models are concerned, the S.A. formulation is needlessly general in one sense (in that the linearity of the model is ignored) and restrictive in another (because of the assumption of conditionally independent observations). Farden [15] has recently proved that linear regression equations can be solved by stochastic approximation methods even if the observations are correlated, provided some mild constraints are imposed on their fourth moments. Even this result is incomplete, since little is known about the convergence rate of the algorithm.

Since we are considering linear channel models, it seems reasonable to try to extend classical linear least squares estimation methods to time varying problems. We shall take this approach in the succeeding sections.

The organization of this chapter is as follows. In section II, we discuss some well known methods of time invariant parameter estimation, such as Gauss least squares, Kalman sequential identification and stochastic approximation. Section III contains the development of an exponentially weighted least squares algorithm. Section IV is devoted to an analysis of the algorithm. The "misadjustment" noise, the error in tracking

time varying parameters, errors due to additive noise and the choice of input signals for application is doppler distorted acoustic communications channel are discussed. Section IV.5. develops an approximation to the proposed algorithm, which we call the "Approximate Least Mean Squares" algorithm (ALMS). We conclude with a summary of our work, some possibilities for future research and a list of references.

The assumptions on the input signal are that

1. $\lim_{n\to\infty} \frac{1}{n} \sum_{k}^{n} \underline{s}_k^* \, \underline{s}_k^T \to \underline{R}_{ss}$.

The limit may be in quadratic mean or an almost sure sense for stochastic inputs.

2. $\lim_{n\to\infty} \frac{1}{n} \sum_{k}^{n} |s_k(1)|^2 \, |s_k(m)|^2 < \infty$, $1,m = 1,2,\ldots L$.

This is to ensure that the result of low pass filtering the matrix $\{\underline{s}_k^* \, \underline{s}_k^T\}$ will have bounded average power.

3. $\underline{R}_{ss}$ is positive definite and Hermiltian. In this case, there is a matrix $\underline{M}$ such that $\underline{M}^{*T} = \underline{M}^{-1}$ and $\underline{M}\,\underline{\Lambda}\,\underline{M}^{-1} = \underline{R}_{ss}$; $\underline{\Lambda}$ is diagonal.

We define the Fourier transform of a matrix $\underline{F}$ as the term by term Fourier transform of each element of $\underline{F}$.

i.e. the $(k,l)^{th}$ element of the Fourier transform of $\underline{F}$ is the Fourier transform of the $(k,l)^{th}$ element of $\underline{F}$.

Section II.1.

Time invariant least squares algorithms.

1. Classical Least squares algorithms.

If the parameters $\underline{c}$ in eqn. (1.2) are time invariant, the least squares estimate of $\underline{c}$ at time n is well known [57] and is given by

$$\hat{\underline{c}}_n = \left(\sum_1^n \underline{s}_k^* \underline{s}_k^T\right)^{-1} \left(\sum_1^n \underline{s}_k^* x_k\right) \qquad (2.1.1)$$

This can be expressed equivalently as

$$\hat{\underline{c}}_n = \left(\frac{1}{n}\sum \underline{s}_k^* \underline{s}_k^T\right)^{-1} \left(\frac{1}{n}\sum \underline{s}_k^* x_k\right)$$

Defining $\underline{P}_n = \left(\frac{1}{n}\sum \underline{s}_k^* \underline{s}_k^T\right)^{-1}$, $\hat{\underline{c}}_n$ can be expressed in the form of an iterative algorithm [ ]:

$$\hat{c}_{n+1} = \hat{c}_n + \frac{\underline{P}_n \underline{s}_{n+1}^* (x_{n+1} - \underline{s}_{n+1}^T \hat{\underline{c}}_n)}{n + \underline{s}_{n+1}^T \underline{P}_n \underline{s}_{n+1}^*}$$

$$\underline{P}_{n+1} = \frac{n+1}{n}\left[\underline{P}_n - \frac{\underline{P}_n \underline{s}_{n+1}^* \underline{s}_{n+1}^T \underline{P}_n}{n + \underline{s}_{n+1}^T \underline{P}_n \underline{s}_{n+1}^*}\right] \qquad (2.1.2)$$

Since $\lim_{n\to\infty} \frac{1}{n} \underline{s}_k \underline{s}_k^{*T} \to \underline{R}_{ss}$ and $\lim_{n\to\infty} \frac{1}{n} \underline{s}_k^* x_k \to \underline{R}_{sx}$ almost surely and in quadratic mean, it is easy to show that $\hat{\underline{c}}_n \to c$ almost surely and in quadratic mean [10,58]. It is not quite clear what advantages, if any, the recursive formulation (2.1.2) has over (2.1.1)

Excellent discussions of linear least squares estimates of time invariant parameters can be found in most texts on statistical inference, e.g. [57].

Section II.2

A. Relationship of MMSE estimation with Kalman Filtering.

Assuming the same model as in the previous subsection,

$$x_k = \underline{s}_k^T \underline{c} + w_k \qquad (2.2.1)$$

We wish to find the linear least squares estimate of $\underline{c}$, i.e. the value of $\underline{c}$ that minimizes

$$E\{(x_k - \underline{s}_k^T \hat{\underline{c}})^T (x_k - \underline{s}_k^T \hat{\underline{c}})\} \qquad (2.2.2)$$

This problem can be cast in a form to which the sequential state estimation algorithms due to Kalman and Bucy [30,31] can be applied. An extensive discussion of Kalman Filtering methods can be found, among other sources, in [3], [8], [17], [28], [30], [31], [38], [39], [43]. The following discussion is based on the work of Brown [6], Chien and Fu [7], de Figureido [12] and Jones [27].

The "state equation" for $\underline{c}$ can be written as

$$\underline{c}_{k+1} = \underline{c}_k \qquad (2.2.3)$$

$$x_k = \underline{s}_k^T \underline{c}_k + w_k \qquad (2.2.4)$$

The trivial equation (2.2.3) simply expresses the fact the $\underline{c}$ is time invariant.

The algorithm for the sequential updating of the estimate

of $\hat{\underline{c}}_n$ can be easily derived (see the cited references) and is known to be

$$\hat{\underline{c}}_{n+1} = \hat{\underline{c}}_n + \underline{P}_n \underline{s}_{n+1}[\underline{s}_{n+1}^T \underline{P}_n \underline{s}_{n+1} + \sigma^2]^{-1} - (x_{n+1} - \underline{s}_{n+1}^T \hat{\underline{c}}_n) \qquad (2.2.5)$$

$$\underline{P}_{n+1} = \underline{P}_n - \underline{P}_n \underline{s}_{n+1} [\underline{s}_{n+1}^T \underline{P}_n \underline{s}_{n+1} + \sigma^2]^{-1} \underline{s}_{n+1}^T \underline{P}_n$$

($\hat{\underline{c}}_o$ and $\hat{\underline{P}}_o$ arbitrary)

If we replace $\underline{P}_n$ by $\frac{\underline{P}_n}{\sigma^2}$, this algorithm is identical to the adaptive sequential least squares algcrithm of the previous subsection (equations (2.1.2.)).

B. Extension to Nonlinear Regression Problems.

If the relation between the observations $\{x_k\}$ and the unknown parameters $\{\underline{s}_k\}$ is nonlinear, the method of quasilinearization can be used to derive a sequential adaptive algorithm. In this case, the Kalman estimator model is

$$\underline{c}_{n+1} = \underline{c}_n \qquad (2.2.3)$$

$$x_n = f_n(\underline{s}_n, \underline{c}_n) + w_n \qquad (2.2.6)$$

where $\{f_n(\underline{s}_n, \underline{c}_n)\}$ are known functions of $\underline{s}_n$ and $\underline{c}_n$. It is assumed that $f_n(., \underline{c}_n)$ is continuously differentiable with respect to $\underline{c}_n$.

Expanding (2.2.6) in a Taylor series about the latest (and presumably, the best) available estaimte of $\underline{c}$, viz. $\hat{\underline{c}}_{n-1}$, and re-

taining only the first two terms

$$x_n \simeq f_n(\underline{s}_n, \hat{\underline{c}}_{n-1}) + \frac{\partial f_n}{\partial \underline{c}_n}^T (\underline{s}_n, \hat{\underline{c}}_{n-1}) \cdot (\underline{c} - \hat{\underline{c}}_{n-1}) + w_n \quad (2.2.7)$$

Define

$$\underline{H}_n = \frac{\partial f_n}{\partial \underline{c}_n} (\underline{s}_n, \hat{\underline{c}}_{n-1})$$

Equation (2.2.6) can be written approximately as

$$x_n - f_n(\underline{s}_n, \hat{\underline{c}}_{n-1}) + \underline{H}_n^T \hat{\underline{c}}_{n-1} = \underline{H}_n^T \underline{c} + w_n \quad (2.2.8)$$

Equations (2.2.3) and (2.2.8) are in a form to which the results of Sec. II.2.A can be applied. The equations for the sequential estimation of $\underline{c}$, similar to equations (2.2.5), turn out to be

$$\hat{\underline{c}}_{n+1} = \hat{\underline{c}}_n + \underline{P}_n \underline{H}_{n+1} (\underline{H}_{n+1}^T \underline{P}_n \underline{H}_{n+1} + \sigma^2)^{-1} (x_{n+1} - f_{n+1}(\underline{s}_{n+1}, \hat{\underline{c}}_n))$$

$$\underline{P}_{n+1} = \underline{P}_n - \underline{P}_n \underline{H}_{n+1} (\underline{H}_{n+1}^T \underline{P}_n H_{n+1} + \sigma^2)^{-1} \underline{H}_{n+1} \underline{P}_n \quad (2.2.9)$$

Convergence of this algorithm is hard to prove. Following Brown [6], it appears that the algorithm will convergence if $f_k(\cdot, \cdot)$ is well behaved in the sense that it satisfies certain Lipshitz conditions.

C. Extension to time varying regression analysis.

The state equations that describe the "evolution" of $\underline{c}$ in the last two subsections assume that $\underline{c}$ is time invariant. A

possible extension of equation (2.2.3) is

$$\underline{c}_{n+1} = g_n(\underline{c}_n) + w_{1n} \qquad (2.2.10)$$

where w_{1n} is a sequence of stochastic forcing functions whose covariance is known. It is assumed that w_{1n} is uncorrelated with w_n in eq. (2.2.4). $\{g_n(\cdot)\}$ is a known sequence of functions. Kalman filtering equations can be applied to this model to update $\{\underline{c}_n\}$ sequentially. Since the assumption that the exact functional form of the evolution of $\{\underline{c}_n\}$ is known appears to be very restrictive, we shall not give details of the derivation here.

Section II.3.

Stochastic Approximation.

One version of the stochastic approximation algorithms for the estimation of the time invariant parameters $\underline{c}$ in eq. (2.2.1) is of the recursive form

$$\hat{c}_{n+1} = \hat{c}_n + \alpha_n (x_n - \underline{s}_n^T \hat{\underline{c}}_n)\underline{s}_n \qquad (2.3.1)$$

where α_n is a time varying gain sequence such that $\sum^{\infty} \alpha_n = \infty$, $\lim_{n\to\infty} \alpha_n \to 0$. A popular choice of α_n is $\alpha_n = 1/n$.

Under some weak conditions on the fourth cumulants of x_k and $\underline{s}_k$, Farden [15] has shown that $\lim_{n\to\infty} \hat{\underline{c}}_n \to \underline{c}$ almost surely. A considerable body of literature exists on the theory and applications of stochastic approximation methods for the solution of nonlinear regression equations [1,3,11,13,15,17,21,29,32,44,50, to list a few].

The main advantages of equation (2.3.1) are that it requires relatively few storage locations and multiplications per iteration. Its limitations are that its convergence rate is usually slow, it cannot track time variations in $\underline{c}$ and its error propagation behavior is unknown. Moreoever, for linear channel models, other efficient, well understood algorithms for the estimation of $\underline{c}$ exist (eq.(2.1.1)). In view of these objections, we omit more detailed discussion of stochastic approximation algorithms.

Section. III.

Sequential least squares algorithms for approximate estimation of slowly time varying regression parameters.

III-1. Introduction

In the previous section, it was assumed that the linear channel to be identified was time invariant. This assumption is not very realistic when modeling underwater acoustic communication channels.

In this section, we permit $\underline{c}$ to be time varying, while still retaining the assumption that the channel is linear. We consider models of the form

$$x_n = \underline{s}_n^T \underline{c}_n + w_n \qquad (3.1.1)$$

and the problem is to identify the time varying parameters $\underline{c}_n$ given $\{x_n\}$ and $\{\underline{s}_n\}$. All the variables in eq. (3.1.1) are assumed to be complex.

If an ensemble $\{\underline{s}_n^i\}_{i=1}^I$ of inputs and $\{x_n^i\}_{i=1}^I$ of outputs is available at each time instant n, $\{\underline{c}_n\}$ can be estimated, at least in principle, by formulating, the classical regression problem

$$x_n^i = \underline{s}_n^{iT} \underline{c}_n + w_n^i, \quad i = 1,2,\ldots I. \qquad (3.1.2)$$

It may be feasible to transmit only one input at every instant of time, i.e. I=1 in equation (3.1.2). In this case, only one equation is available to solve for the L components of

$\underline{c}_n$ for each n. Unless some structure is imposed on the nature of the functions $\underline{c}_n$, the problem is unsolvable.

One possibility is to represent $\underline{c}_n$ over a finite interval of time $n_1 \leq n \leq n_2$ as a linear combination of small number of known basis functions. $\underline{c}_n$ is modeled as

$$\underline{c}_n = \sum_{i=1}^{I} \underline{c}(i)\ \phi_n(i) \qquad (3.1.3)$$

Now the problem of estimating $\underline{c}_n$ over the time interval $n_1 \leq n \leq n_2$ reduces to that of estimating the (IL) time invariant parameters $\underline{c}(i)$, i = 1, ...I.

The simplest choice of $\{\phi_n(i)\}$ in eq. (3.1.3) is to take $\phi_n(0) = 1$, $\phi_n(i) = 0$ for all other i, i.e., $\underline{c}_n$ is assumed to be stepwise constant. More general basis functions, such as polynomials and trigonometric functions, have been familiar to time series analysts for a long time. Models of the form (3.1.3) have proved to be useful in many engineering applications, such as linear predictive analysis of speech [36], system identification [34] and adaptive antenna arrays [41].

Markov modeling of $\underline{c}_n$ and the associated sequential "state" estimation algorithm have been discussed in the previous section.

Quite often, nothing is known about the structure of $\underline{c}_n$ except that it is slowly time varying. In the rest of this section, we shall consider sequential least squares algorithms that give approxiamte estimates of $\underline{c}_n$.

Section III.2.

An Exponentially weighted Least squares Algorithm.

The model (3.1.1) is assumed. If $\underline{c}_n = \underline{c}$, a time invariant vector, the linear minimum mean squares estimate of $\underline{c}$ given $\{x_n\}$ and $\{s_n\}$is

$$\hat{\underline{c}}_n = \left(\frac{1}{n}\sum_{k=1}^{n} \underline{s}_k^* \underline{s}_k^T\right)^{-1} \left(\sum_k \frac{1}{n} x_k \underline{s}_k^*\right) \qquad (3.2.1)$$

The derivation leading to this estimate treats all the inputs, $\{s_k\}_{k=1}^{n}$ and observations, $\{x_k\}$, with the same importance. If $\underline{c}$ were slowly time varying, however, it seems more reasonable to attach decreasing importance to past inputs and observations. The idea is not new; Jones [27] and Harris [25] have used a similar technique for the updating of autoregressive model parameter estimates.

We replace $\{\frac{1}{n} \underline{s}_k^* \underline{s}_k^T\}$ and $\{\frac{1}{n} \underline{s}_k^* x_k\}$ in equation (3.2.1) by

$\{\underline{W}_{n,k} \underline{s}_k^* \underline{s}_k^T\}$ and $\{\underline{W}_{n,k} \underline{x}_k \underline{s}_k^*\}$ respectively. $\{\underline{W}_{n,k}\}$ is a sequence of weight matrices such that

$$\sum_{k=1}^{n} \underline{W}_{n,k} = \underline{I}$$

$$\underline{W}_{n,k} = \underline{0} \text{ for } n > k$$

$$\underline{W}_{n,k} \to \underline{0} \text{ as } n - k \to \infty.$$

The only sequence $\{W_{n,k}\}$ that we have found useful for the purpose of sequential least squares algorithms is of the form

$$\underline{\underline{W}}_{n,k} = (\underline{\underline{I}}-\underline{\underline{Q}}^n)^{-1}\ (\underline{\underline{I}}-\underline{\underline{Q}})\cdot\underline{\underline{Q}}^{n-k} \tag{3.2.2}$$

where the matrix $\underline{\underline{Q}}$ has all its eigenvalues in the open interval $(0,1)$.

With this modification, equation (3.2.1) is replaced by

$$\underline{c}_n = (\underline{\underline{I}}-\underline{\underline{Q}}^n)^{-1}\ (\underline{\underline{I}}-\underline{\underline{Q}})\ \sum_{k=1}^{n}\ \underline{\underline{Q}}^{n-k}\ \underline{s}_k^*\ \underline{s}_k^T)^{-1}\cdot$$

$$((\underline{\underline{I}}-\underline{\underline{Q}}^n)^{-1}(\underline{\underline{I}}-\underline{\underline{Q}})\ \sum_{k=1}^{n}\underline{\underline{Q}}^{n-k}\ x_k\ \underline{s}_k^*) \tag{3.2.3}$$

Let

$$\underline{\underline{P}}_n = ((\underline{\underline{I}}-\underline{\underline{Q}}^n)^{-1}\ (\underline{\underline{I}}-\underline{\underline{Q}})\ \sum_{k=1}^{n}\underline{\underline{Q}}^{n-k}\ \underline{s}_k^*\ \underline{s}_k^T)^{-1} \tag{3.2.4}$$

and

$$\underline{d}_n = (\underline{\underline{I}}-\underline{\underline{Q}}^n)^{-1}\ (\underline{\underline{I}}-\underline{\underline{Q}})\ \sum_{k=1}^{n}\ \underline{\underline{Q}}^{n-k}\ x_k\ \underline{s}_k^* \tag{3.2.5}$$

Then

$$\underline{\underline{P}}_{n+1} = \{(\underline{\underline{I}}-\underline{\underline{Q}}^{n+1})^{-1}\ (\underline{\underline{I}}-\underline{\underline{Q}})\cdot\underline{\underline{Q}}\cdot\ \sum_{k=1}^{n}\underline{\underline{Q}}^{n-k}\ \underline{s}_k^*\underline{s}_k^T +$$

$$(\underline{\underline{I}}-\underline{\underline{Q}}^{n+1})^{-1}\ (\underline{\underline{I}}-\underline{\underline{Q}})\ \underline{s}_{n+1}^*\ \underline{s}_{n+1}^T\}^{-1}$$

$$= \{(\underline{\underline{I}}-\underline{\underline{Q}}^{n+1})^{-1}\ (\underline{\underline{I}}-\underline{\underline{Q}}^n)\underline{\underline{Q}}\ \underline{\underline{P}}_n^{-1} +$$

$$(\underline{\underline{I}}-\underline{\underline{Q}}^{n+1})^{-1}\ (\underline{\underline{I}}-\underline{\underline{Q}})\ \underline{s}_{n+1}^*\ \underline{s}_{n+1}^T\}^{-1} \tag{3.2.6}$$

We now make use of the matrix inversion lemma [43]

$$(\underline{I}+\underline{a}\,\underline{b}^T)^{-1} = \underline{I} - \frac{\underline{a}\,\underline{b}^T}{1 + \underline{b}^T\underline{a}} \qquad (4.19) \quad (3.2.7)$$

Applying this result to (3.2.6)

$$\underline{P}_{n+1} = \left\{ \underline{I} - \frac{\underline{P}_n\,\underline{Q}^{-1}\,(\underline{I}-\underline{Q}^n)^{-1}\,(\underline{I}-\underline{Q})\,\underline{s}_{n+1}^*\,\underline{s}_{n+1}^T}{1 + \underline{s}_{n+1}^T\,\underline{P}_n\,\underline{Q}^{-1}\,(\underline{I}-\underline{Q}^n)^{-1}(\underline{I}-\underline{Q})\underline{s}_{n+1}^*} \right\}$$

$$\underline{P}_n\,\underline{Q}^{-1}\,(\underline{I}-\underline{Q}^n)^{-1}\,(\underline{I}-\underline{Q}^{n+1}) \qquad (3.2.8)$$

$$\underline{d}_{n+1} = (\underline{I}-\underline{Q}^{n+1})^{-1}\,(\underline{I}-\underline{Q}^n)\cdot\underline{Q}\,\underline{d}_n + (\underline{I}-\underline{Q}^{n+1})^{-1}\,(\underline{I}-\underline{Q})\,\underline{x}_{n+1}\,\underline{s}_{n+1}^* \qquad (3.2.9)$$

Then $\hat{\underline{c}}_{n+1} = \underline{P}_{n+1}\,\underline{d}_{n+1}$ can be calculated to be

$$\underline{c}_{n+1} = \underline{c}_n + \frac{\{\underline{P}_n\,\underline{Q}^{-1}(\underline{I}-\underline{Q}^n)^{-1}(I-Q)(x_{n+1}-\underline{s}_{n+1}^T\underline{c}_n)\}\cdot\underline{s}_{n+1}^*}{\{1 + \underline{s}_{n+1}^T\,\underline{P}_n\,\underline{Q}^{-1}\,(\underline{I}-\underline{Q}^n)^{-1}(\underline{I}-\underline{Q})\underline{s}_{n+1}^*\}} \qquad (3.2.10)$$

Equations (3.2.8) and (3.2.10) constitute the exponentially weighted least squares (EWLS) algorithm. The term $(x_{n+1}-\underline{s}_{n+1}^T\,\hat{\underline{c}}_n)$ is the prediction error at time n+1. Since all the eigenvalues of $\underline{Q}$ are in (0,1), $\underline{Q}^n \to \underline{0}$ as $n\to\infty$ in norm. Using this fact, the asymptotic EWLS algorithm can be written as

$$\hat{\underline{c}}_{n+1} = \hat{\underline{c}}_n + \frac{\underline{P}_n\,\underline{Q}^{-1}\,(\underline{I}-\underline{Q})(x_{n+1}-\underline{s}_{n+1}^T\,\hat{\underline{c}}_n)s_{n+1}^*}{1 + \underline{s}_{n+1}^T\,\underline{P}_n\,\underline{Q}^{-1}\,(\underline{I}-\underline{Q})\,\underline{s}_{n+1}^*}$$

(4.20)

$$\underline{P}_{n+1} = \underline{P}_n \underline{Q}^{-1} - \frac{\underline{P}_n \underline{Q}^{-1} (\underline{I}-\underline{Q}) \underline{s}_{n+1}{}^* \underline{s}_{n+1}{}^T}{1+\underline{s}_{n+1} \underline{P}_n \underline{Q}^{-1} (\underline{I}-\underline{Q}) \underline{s}_{n+1}{}^T} \qquad (3.2.11)$$

$n = 0,1,2,\ldots;$ $\underline{P}_o$ arbitrary

As $\underline{Q} \to \underline{I}$, this algorithm becomes identical to the time invariant least squares algorithm.

Section IV.

Analysis of the EWLS algorithm.

This section is devoted to the study of the EWLS algorithms for two $\underline{Q}$ matrices:

$$\underline{Q} = \underline{Q}_{(1)} = q\,\underline{I}\,, \quad 0 < q < 1 \quad , \text{ and}$$

$$\underline{Q} = \underline{Q}_{(2)} = (\underline{I} + \alpha\,\underline{R}_{ss})^{-1}\,\underline{M}\,[(1+\underline{\alpha}\lambda_i)^{-1}]\underline{M}^{-1}$$

$\underline{Q}_{(1)}$ is the simplest choice of $\underline{Q}$. The utility of $\underline{Q}_{(2)}$ is in deriving a computationally simpler Approximate Least Mean Squares (ALMS) algorithm (Section IV.5).

Since the eigenvalues of $\underline{Q}$ must lie between 0 and 1, it is clear that

$0 < \alpha < \frac{1}{\lambda_{max}}$, where λ_{max} is the largest eigenvalue of $\underline{R}_{ss}$. This is a necessary, but not sufficient condition for the stability of the EWLS algorithm.

Section IV.1.

Properties of $\underline{P}_n^{-1}$

$\underline{P}_n^{-1}$ was defined in Sec. III (eq. 3.2.4) for large values of n as

$$\underline{P}_n^{-1} = (\underline{I}-\underline{Q}) \sum_{k=1}^{n} \underline{Q}^{n-k}\,\underline{s}_k^*\,\underline{s}_k^T \qquad (4.1.1)$$

The key to the analysis of $\underline{P}_n^{-1}$ is to recognize that $\underline{P}_n^{-1}$ is the output of a matrix valued low pass filter $\{\underline{Q}^k\}$ whose input is $\{\underline{s}_k^*\,\underline{s}_k^T\}$. Therefore the mean value of $\underline{P}_n^{-1}$ is $\underline{R}_{ss}$,

regardless of the choice of $\underline{Q}$ and the nature of the signal.

The variable component of $\underline{P}_n^{-1}$ is harder to determine and is strongly dependent on the exact structure of the signal and the choice of $\underline{Q}$. We shall consider the following cases.

1. $\underline{Q} = \underline{Q}_{(1)}$; and $\{\underline{s}_k^* \underline{s}_k^T\}$ has a line spectrum at frequencies f(1), f(2) . . f(m). This will be the case, for example, if $\{\underline{s}_k\}$ is a periodic pulse train or a sum of sinusoids. In this case, the $(i,1)^{th}$ element of $\underline{P}_n^{-1}$ is the form

$$\underline{P}_n^{-1}(i,1) = \sum_m g(i,1)(m)\, e^{-j2\pi f(m)\cdot n}$$

and

$$\underline{P}_n^{-1}(i,1) - \bar{\underline{P}}_n^{-1}(i,1) =$$

$$(1-q)\cdot \sum_m \frac{1}{1 - q\, e^{-j\pi f(m)}}\, g(i,1)(m).$$

similarly, for $\underline{Q} = \underline{Q}_2$,

$[\underline{P}_n^{-1}(k,1) - \bar{\underline{P}}_n^{-1}(k,1)]$ has the form

$$\sum_{i=1}^{L} \sum_m \frac{\lambda_i \alpha}{1+\lambda_i \alpha - e^{-j2\pi f(m)}}\, g'(k,1)\,(f(m))$$

where $g'(k,1)(f(m))$ is a bounded linear combination of the elements of $\underline{s}_k^* \underline{s}_k^T$.

These expressions will be uniformly small if each f(m) is not within the passbands of the low filters $\{\underline{Q}^k\}$. If a periodic

pulse train is used as input signal, it is necessary that its pulse repetition period be much smaller than the time constants of the averaging filter, $\underline{Q}$.

2. $\underline{Q} = \underline{Q}_{(1)}$; $\{\underline{s}_k^* \underline{s}_k^T\}$ has a continuous bounded spectrum. An example of this is the pseudorandom noise sequence. In this case, the average power (or variance) of the variable component of each element of $\underline{P}_n^{-1}$ is

$$\text{var}\{(\underline{P}_n^{-1} - \underline{\bar{P}}_n^{-1})\ (i,1)\} =$$

$$\int_{-1/2}^{1/2} \frac{(1-q)^2}{|1-q\ e^{-j2\pi f}|^2} |\underline{s}_k^*(i)\ \underline{s}_k(1)(f)|^2 \quad (4.1.3)$$

where f indicates that the integral is evaluated after omitting the spectral line at f = 0.

If $g''(i,1) = \max_f |s_k^*(i)\ s_k(1)(f)|^2 < \infty$,

then

$$\text{Var}\{(P_n^{-1} - P_n^{-1})\}(i,1) \leq (1-q) \cdot g''(i,1)$$

so that for small (1-q), the time varying component of $\underline{P}_n^{-1}$ has small power.

Similarly, for $\underline{Q} = \underline{Q}_2$,

$\text{Var}\{(\underline{P}_n^{-1} - \underline{P}_n^{-1})(i,1)\} \leq \alpha\ b\ (i,1)$, where b(i.1) is some bounded constant.

Stability of $\underline{P}_n$

Since $\underline{P}_n^{-1}$ is a time varying matrix, there is little guarantee that it is invertible at all. We can reasonably hope that if the sum of the mean power of the variable components of all the elements of $\underline{P}_n^{-1}$ is much smaller than the smallest eigenvalue of its mean, viz. $\underline{R}_{ss}$, then $\underline{P}_n^{-1}$ would be invertible. Due to the complexity and signal dependence of the expressions for the time varying components of $\underline{P}_n^{-1}$, it is hard to make any general statements about the conditions for the stability of the EWLS algorithm.

As an illustration, assume that the input signals are independent gaussian random variables such that $\underline{R}_{ss} = \lambda \underline{I}$. In this case, with $\underline{Q} = \underline{Q}_{(1)}$,

$$\underline{\bar{P}}_n^{-1} = \lambda \underline{I}$$

$$\text{Var}\{\underline{P}_n^{-1}(k,l)\} = \lambda^2 \frac{(1-q)}{1-q} \quad \text{for } k \neq l$$

$$= 2\lambda^2 \frac{(1-q)}{1-q} \quad \text{for } k = l.$$

In order to obtain adequate stability, we set

$$\sum_{k,l} \text{Var}\{P_n^{-1}(k,l)\} \leq \lambda_{min}^2 = \lambda^2$$

i.e. $(L+L^2)(1-q) \leq 1 + q$ or

$$(1-q) \leq \frac{1+q}{L^2+L} \simeq \frac{2}{L^2+L}$$

with $\underline{Q} = \underline{Q}_{(2)}$, $\alpha \leq \frac{2}{(L^2+L)\lambda}$.

Anticipating the results of Section IV.5., this is our condition on α so that the LMS algorithm is stable, with independent gaussian inputs. A more accurate analysis by Merriam [39] indicates that the LMS algorithm will be stable if

$\alpha \leq \frac{2}{L+2}$. Therefore, our criterion for stability is rather pessimistic.

We shall denote $\underline{P}_n^{-1}$ by $\bar{\underline{P}}_n^{-1} + \underline{N}_u$, where $\underline{N}_u$ lumps togetehr all the time varying components of $\underline{P}_n^{-1}$. Hopefully, each element of $\underline{N}_u$ has small variance, and therefore we are justified in making the approximation

$$\underline{P}_n = (\underline{P}_n^{-1})^{-1} \simeq \underline{R}_{ss}^{-1} - \underline{R}_{ss}^{-1} \underline{N}_u \underline{R}_{ss}^{-1}$$

Section IV.2.

Tracking Behavior of the EWLS algorithm

The asymptotic expression for the estimate of $\underline{c}_n$ at time n (eq.(3.2.3)) can be written as

$$\hat{\underline{c}}_n = ((\underline{\underline{I}}-\underline{\underline{Q}}) \sum_{k=1}^{n} \underline{\underline{Q}}^{n-k} \underline{s}_k^* \underline{s}_k^T)^{-1} (\underline{\underline{I}}-\underline{\underline{Q}}) \sum_{k=1}^{n} \underline{\underline{Q}}^{n-k} \underline{s}_k^* (\underline{s}_k^T \underline{c}_k + w_k) \quad (4.2.1)$$

If $\underline{c}_n = \underline{c}$, a constant, $\hat{\underline{c}}_n$ is an unbiased estimate of $\underline{c}$. Since $\hat{\underline{c}}_n$ is linear in $\underline{c}_n$ and w_n, the effect of additive noise will be considered separately from that of time variation of $\underline{c}_n$. In the rest of this subsection, w_n will be assumed to be zero.

A step variation in $\underline{c}_n$ is easily analyzable. Let $\underline{c}_n = \underline{c}_{(1)}$ for $n<n_1$ and $\underline{c}_n = \underline{c}_{(2)}$ for $n \geq n_1$. Then at time n_1+n,

$$\hat{\underline{c}}_{n+n_1} = \underline{\underline{P}}_{n+n_1} \{(\underline{\underline{I}}-\underline{\underline{Q}}) \sum_{k=1}^{w_1+n} \underline{\underline{Q}}^{n_1+n-k} \underline{s}_k^* \underline{s}_k^T \underline{c}_{(2)} +$$

$$(\underline{\underline{I}}-\underline{\underline{Q}}) \underline{\underline{Q}}^n \sum_{k=1}^{n_1} \underline{\underline{Q}}^{n-k} \underline{s}_k^* \underline{s}_k^T\} (c_{(1)}-c_{(2)}) \quad (4.2.2)$$

The first term of (4.2.2) is recognized as $c_{(2)}$. The second term, which represents the error in the estiamte of $\underline{c}_{n_1}$ approximately equal to $\underline{\underline{Q}}^n(c_{(1)}-c_{(2)})$ and approaches zero at an exponential rate.

Next we consider the case

$$\underline{c}_n = \underline{c}_{(1)} e^{j2\pi f_o r}$$

Now

(4.27)

$$\hat{\underline{c}}_n = \underline{P}_n (\underline{I}-\underline{Q}) \sum_{k=1}^{n} (\underline{Q}\, e^{-j2\pi f_o})^{n-k} \underline{s}_k^* \underline{s}_k^T \underline{c}_{(1)}\, e^{j2\pi f_o} \qquad (4.2.3)$$

The weights $\{(\underline{Q}\, e^{-j2\pi f_o})^k\}$ represents the impulse response to a first order bandpass filter whose freuqency is at $-f_o$. This bandpass filter operates on the sequence $\{\underline{s}_k^* \underline{s}_k^T\}$ and the means value of the output of the filter $\simeq (\underline{I}-\underline{Q}\, e^{-j2\pi f_o})^{-1} \underline{R}_{ss}$. The variable component of the output, denoted by $\underline{N}_u'$, is highly dependent on the detailed signal structures and can be treated in exactly the same manner as in the analysis of $\underline{P}_n^{-1}$. (Sec.IV.1).

$$\hat{\underline{c}}_n \simeq (\underline{R}_{ss}^{-1} + \underline{R}_{ss}^{-1} \underline{N}_u \underline{R}_{ss}^{-1})\ (\underline{I}-\underline{Q})\ (\underline{I}-\underline{Q}\, e^{-j2\pi f_o})^{-1}$$

$$(\underline{R}_{ss} + \underline{N}_u')\ \underline{c}_{(1)}\, e^{-j2\pi f_o}$$

$$\simeq \underline{R}_{ss}^{-1} (\underline{I}-\underline{Q})\ (\underline{I}-\underline{Q}\, e^{-j2\pi f_o})^{-1} \underline{R}_{ss}$$

$$- \underline{R}_{ss}^{-1} \underline{N}_u \underline{R}_{ss}^{-1} (\underline{I}-\underline{Q})\ (\underline{I}-\underline{Q}\, e^{-j2\pi f_o})^{-1} \underline{R}_{ss}$$

$$+ \underline{R}_{ss}^{-1} (\underline{I}-\underline{Q})\ (\underline{I}-\underline{Q}\, e^{-j2\pi f_o})^{-1} \underline{Nu}'\}.\ \underline{c}_n$$

For $\underline{Q} = q\underline{I}$,

$$\hat{\underline{c}}_n \simeq \frac{1-q}{1-q\, e^{-j2\pi f_o}} \underline{c}_n + \frac{1-q}{1-q\, e^{-j2\pi f_o}} \underline{R}_{ss}^{-1} (\underline{N}_u'-\underline{N}_u)\underline{c}_n \qquad (4.2.4)$$

and for $\underline{Q} = (\underline{I}+\alpha \underline{R}_{ss})^{-1}$,

(4.28)

$$\hat{\underline{c}}_n = \underline{\underline{M}} \left[\frac{\lambda_i \alpha}{1 + \lambda_i \alpha - e^{-j2\pi f_o}}\right] \underline{\underline{M}}^{-1} \underline{c}_n$$

$$- \underline{\underline{R}}_{ss}^{-1} \underline{\underline{N}}_u \underline{\underline{M}} \left[\frac{\lambda_i \alpha}{1 + \lambda_i \alpha - e^{-j2\pi f_o}}\right] \underline{\underline{M}}^{-1} \underline{c}_n$$

$$+ \underline{\underline{M}} \left[\frac{\alpha}{1 + \lambda_i \alpha - e^{-j2\pi f_o}}\right] \underline{\underline{M}}^{-1} \underline{\underline{N}}_u' \underline{c}_n \qquad (4.2.5)$$

We showed that each term of $\underline{\underline{N}}_u$ (and similarly $\underline{\underline{N}}_u'$) is a time varying function whose variance (or power, in the case of deterministic signals) is proporational to (1-q) or α, according as $\underline{\underline{Q}} = q\,\underline{\underline{I}}$ or $\underline{\underline{Q}} = (\underline{\underline{I}} + \alpha \underline{\underline{R}}_{ss})^{-1}$.

Since (1-q) and α are assuemd to be small, the effects of $\underline{\underline{N}}_u'$ on $\underline{c}_n$ are second order effects compound to the first terms of (4.2.4) and (4.2.5). Further, in order to ensure that $\underline{c}_n \simeq \underline{c}_n$, it is essential that

$$\frac{1 - q}{1 - q\, e^{-j2\pi f_o}} \simeq 1 \quad \text{and} \quad \frac{\lambda_i \alpha}{1 + \lambda_i \alpha - e^{-j2\pi f_o}} \simeq 1.$$

Under this condition, $\underline{\underline{N}}_u'$ and $\underline{\underline{N}}_u'$ will be highly correlated, since they are the results of filtering the same sequence $\{s_k^* s_k^T\}$ with two linear filters with highly overlapping passbands. Therefore each element of $\underline{\underline{N}}_u' - \underline{\underline{N}}_u$ will be negligible in comparison with the first terms of (4.2.4) and (4.2.5).

$$\hat{\underline{c}}_n \simeq \frac{1 - q}{1 - q\, e^{-j2\pi f_o}} \underline{c}_n \qquad \text{for } \underline{\underline{Q}} = q\,\underline{\underline{I}} \qquad (4.2.6)$$

$$\hat{\underline{c}}_n \approx \underline{\underline{M}} \left[\frac{\lambda_i \alpha}{1 + \lambda_i \alpha - e^{-j2\pi f_o}} \right] \underline{\underline{M}}^{-1} \underline{c}_n \text{ for } \underline{\underline{Q}} = (\underline{\underline{I}} + \alpha \underline{\underline{R}}_{ss})^{-1} \quad (4.2.7)$$

Since $\hat{\underline{c}}_n$ is linear in $\underline{c}_n$, the equations (4.2.6) and (4.2.7) can be generalized to the case where $\underline{c}_n$ consists of sums of sinusoids, and in the limit, has an arbitrary spectrum $\tilde{\underline{c}}(f)$.

$$\hat{\underline{c}}_n = \int_{-1/2}^{1/2} \frac{1 - q}{1 - q\, e^{-j2\pi f}} \tilde{\underline{c}}(f)\, e^{j2\pi f}\, df \quad \text{for } \underline{\underline{Q}} = \underline{\underline{Q}}_1 \quad (4.2.8)$$

$$\hat{\underline{c}}_n = \int_{-1/2}^{1/2} \underline{\underline{M}} \left[\frac{\lambda_i \alpha}{1 + \lambda_i \alpha - e^{-j2\pi f}} \right] \underline{\underline{M}}^{-1} e^{j2\pi f}\, \tilde{\underline{c}}(f)\, dt \text{ for } \underline{\underline{Q}} = \underline{\underline{Q}}_2 \quad (4.2.9)$$

These expressions for $\hat{\underline{c}}_n$ show that it is highly dependent on the spectrum of $\underline{c}_n$. In addition, for $\underline{\underline{Q}} = \underline{\underline{Q}}_2$, $\hat{\underline{c}}_n$ also depends on the data correlation, via $\underline{\underline{M}}$. If special forms are assumed for $\underline{c}(f)$ and $\underline{\underline{M}}$, it may be possible to evaluate these integrals explicitly. For example, if $\underline{\underline{M}}$ is taken to be the identity matrix and $\underline{c}_n$ is assumed to have a uniform lowpass spectrum, a rational spectrum or a harmonic spectrum (4.2.8) and (4.2.9) may be evaluated explicitly. We shall not carry out the details, but simply point out that if the bandwidth of each component of $\tilde{\underline{c}}(f)$ is much smaller than the bandwidth of the matrix low pass filter whose transfer function is $(\underline{\underline{I}} - \underline{\underline{Q}})\ (\underline{\underline{I}} - \underline{\underline{Q}}\, e^{-j2\pi f})^{-1}$, $\hat{\underline{c}}_n \simeq \underline{c}_n$.

An important point should be noted if periodic pulse trains are used as input signals. If the input signal has a period T, the spectrum of $\{\underline{s}_k^*\ \underline{s}_k^T\}$ will have harmonic components at frequencies of integral multiples of 1/T. It is essential to ensure that 1/T is much larger than the sum of the bandwidths of the low-

pass filter $\{\underline{Q}^k\}$ and the largest significant frequency component of $\underline{\tilde{c}}(f)$, f_m. Otherwise the bandpass filter $\{(\underline{Q}\ e^{-j2\pi f_o})^k\}$ in eq. (4.2.3) will introduce significant time varying components in $\underline{\hat{c}}_n$.

Section IV.3.

Effect of additive noise.

Let $\{x_k\} = \{w_k\}$ i.e. $\underline{c}_n = \underline{0}$ for all n. Ideally $\hat{\underline{c}}_n$ should be zero for all n. In this section we shall examine how closely $\hat{\underline{c}}_n$ approximates this ideal.

$$\underline{c}_n = \underline{P}_n \, (\underline{I}-\underline{Q}) \sum \underline{Q}^{n-k} \, \underline{s}_k^* \, w_k \qquad (4.3.1)$$

$$E_{w_n} \{\hat{\underline{c}}_n\} = 0 \quad \text{and}$$

$$E_{w_n} \{\hat{\underline{c}}_n \, \hat{\underline{c}}_n^{*T}\} = \underline{P}_n \, (\underline{I}-\underline{Q}) \, (\sum_{k=1}^{w} \underline{Q}^{n-k} \, \underline{s}_k^* \, \sigma^2 \, \underline{s}_k^T (\underline{Q}^{*T})^{n-k}) \, (\underline{I}-\underline{Q}^{*T}) \underline{P}_n^{*T} \qquad (4.3.2)$$

If we take $\underline{Q} = q \, \underline{I}$,

$$E_w \, \hat{\underline{c}}_n \, \hat{\underline{c}}_n^{*T} = (1-q)^2 \, \underline{P}_n \, (\sum_{k=1}^{n} q^{2(n-k)} \, \underline{s}_k^* \, \underline{s}_k^T \, \sigma^2) \, \underline{P}_n$$

$$\simeq (1-q)^2 \, \underline{R}_{ss}^{-1} \, \frac{\underline{R}_{ss}}{(1-q^2)} \, \sigma^2 \, \underline{R}_{ss}^{-1}$$

$$\simeq \frac{\sigma^2(1-q)}{2} \, \underline{R}_{ss}^{-1}$$

This is the mean value of the expectation of $E_w\{\hat{\underline{c}}_n \hat{\underline{c}}_n^{*T}\}$. In addition, this covariance matrix has time varying components with power on the order of $(1-q)^2$, arising from the fact that $\underline{P}_n$ and $\sum_k q^{2(n-k)} \underline{s}_k^* \, \underline{s}_k^T$ have time varying components.

We have not been able to obtain any simplification of (4.3.2) when $\underline{Q} = (\underline{I} + \alpha \underline{R}_{ss})^{-1}$.

Section IV.4.

Choice of input signal for adaptive estimation in doppler corrupted communication channels.

Until now, we have ignored all doppler shift effects and assumed that the observations $\{x_k\}$ can be modeled as a noisy linear combination of the undistorted input signals. In this subsection we consider a more general model of the form

$$x_k = \underline{v}_k^T \underline{c}_k + w_k \qquad (4.4.1)$$

where

$\underline{v}_k$ is defined by

$$\underline{v}_k = \begin{Bmatrix} v_k(1) \\ v_k(2) \\ \cdot \\ \cdot \\ v_k(L) \end{Bmatrix} = \begin{Bmatrix} v(1)(t) \\ v(2)(t) \\ v(3)(t) \\ v(L)(t) \end{Bmatrix}_{t=k} ;$$

$v(i)(t) = s(i)(P(i)\cdot t)$, $i = 1,2,\ldots L$; $P(i)$ = doppler scale associated with the i^{th} component of $\underline{s}(t)$.

In order to focus on the influence of the wideband ambiguity function of $s(t)$ on $\hat{\underline{c}}_n$, we simplify matters by assuming that $\hat{\underline{c}}_n = \underline{c}$ = constant and $w_k = 0$. We have shown in Sections IV.2. and IV.3. how these restrictions can be removed. With these assumptions

$$\hat{\underline{c}}_n = \underline{P}_n \, (\underline{I}-\underline{Q}) \Big(\sum_{k=1}^{n} \underline{Q}^{n-k} \, \underline{s}_k^* \, \underline{v}_k^T\Big)\underline{c} \qquad (4.4.1a)$$

Define

$$\underline{D}_n' = \sum_{k=1}^{n} \underline{Q}^{n-k} \, \underline{s}_k^* \, \underline{v}_k^T \qquad (4.4.2)$$

and

$$\underline{\underline{D}}_n = \sum_{k=1}^{n} \underline{\underline{Q}}^{n-k} \; \underline{s}_k^* \; \underline{s}_k^T \qquad (4.4.3)$$

$\underline{\underline{D}}'_n$ is the output of passing the sequence $\{\underline{s}_k^* \; \underline{v}_k^T\}$ through a lowpass filter with coefficients $\{\underline{\underline{Q}}^k\}$ so that the mean value of $\underline{\underline{D}}'_n$ is the time average of $\{\underline{s}_k^* \; \underline{v}_k^T\}$.

The (hopefully small) variable component of $\underline{\underline{D}}'_n$ will depend on the spectral characteristics of $\{\underline{s}_k^* \; \underline{v}_k^T\}$.

Therefore, the mean value of the $(i,1)^{th}$ element of $\underline{\underline{D}}'_n$ is

$$\underline{\underline{\bar{D}}}'(i,1) = \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} s(i)^*(t)\Big|_{t=k} \cdot v(1)(t)\Big|_{t=k}$$

$$= \lim_{n\to\infty} \frac{1}{n} \sum_{k} s^*(i)(t)\Big|_{t=k} \cdot s(1)(P(1)\cdot t\Big|_{t=k} \qquad (4.4.4)$$

In the absence of doppler scaling,

$$\underline{\underline{\bar{D}}}(i,1) = \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} s^*(i)(t)\Big|_{t=k} \cdot s(1)(t)\Big|_{t=k} \qquad (4.4.5)$$

Define the cross ambiguity function between $s(i)(t)$ and $s(1)(t)$ as:

$$\chi_{s(i),s(1)}(\tau,P) \equiv \lim_{T\to\infty} \frac{1}{T} \int_0^T s^*(i)(t) \; s(1)(P(t-\tau))dt.$$

Assuming that aliasing errors due to sampling are negligible, eq. (4.4.4) is the cross ambiguity function between $s(i)(t)$ and $s(1)(t)$ at zero lag, i.e.

$$\underline{\bar{D}}'(i,1) = \chi_{s(i),s(1)}\ (0,P(1)$$

Similarly,

$$\underline{\bar{D}}(i,1) = \chi_{s(i),s(1)}(0,1)$$

Since $\underline{\hat{c}}_n$ is equal to (from (4.4.1a) and (4.4.2))

$\underline{\hat{c}}_n = \underline{P}_n\ (\underline{I}-\underline{Q}) \cdot \underline{D}'_n\ \underline{c}$, it is clear that the estimation algorithm will be insensitive to doppler scaling if and only if $\underline{\bar{D}}'_n \simeq \underline{\bar{D}}_n$, i.e. $\chi_{s(i),s(1)}(0,P(1))$ is independent of P(1). In other words, for doppler insensitive adaptive estaimtion, s(i) and s(1) should be chosen such that

$$\varkappa_{s(i),s(1)}(0,p(1)) = \varkappa_{s(i),s(1)}(0,1) \qquad (4.4.6)$$

conversely, if $\varkappa_{s(i),s(1)}(o,P(1)) \simeq 0$ for $P(1) \neq 1$, i.e. if s(i) and s(1) are doppler sensitive signals, $\underline{\hat{c}}_n \simeq 0$ irrespective of the value of $\underline{c}$.

A class of wideband doppler insensitive frequency modulated signals has been developed by Altes and Titlebaum [2]. They have the general form $u(t) = a(t)\ e^{jb\ln t}$ where a(t) and b control the time width, bandwidth and rate of variation of instant, aneous frequency of u(t). This class of signals has the property

$$\varkappa_{u,u}(\tau,P) \simeq e^{jb\ln P}\ \varkappa_{uu}(\tau,1).$$

If the input signals $\{\underline{s}_k\}$ are chosen from this class of signals as

$$\{\underline{s}_{(t)}\} = \{a(1)(t)\ e^{jb(1)\ln t},\ a(2)(t)\ e^{jb(2)\ln t},\ \ldots$$

$$a(L)(t)\ e^{jb(t)\ln t}\}^T\ ,$$

the estimate of $\underline{c}$ in the presence of doppler scaling will be

$$\hat{\underline{c}}_n = (c(1)\ e^{jb(1)\ln P(1)}, c(2)^{jb(2)\ln P(2)}, \ldots c(L)\ e^{jb(1)\ln P(L)})^T.$$

While this estaimte is not independent of the doppler scaling factors $(P(1), \ldots P(L))^T$, these appear in $\hat{\underline{c}}_n$ in a relatively tractable fashion as a simple phase shift of each component of $\underline{c}$.

Another class of noise like signals with desirable doppler insensitity properties was suggested to the author by Prof. E. Titlebaum. They are of the form

$$u(t) = \sum_{i=1}^{r} a(i)\ f(P(i)(t))$$

where f(t) is generated by suitably modulating a pseudo random sequence. We have not undertaken any intensive study of the ambiguity functions of such signals.

Section IV.5.

A simplification of the EWLS algorithm.

The EWLS algorithm has been shown to be able to track slow time variations of $\underline{c}_n$ and with a choice of doppler tolerant input signals, to be insensitive to doppler scaling effects. The main practical difficulty in the application of this algorithm is that it is computationally cumbersome. It is worthwhile to look for possible simplification that will make the EWLS algorithm more attractive for real time applications.

Consider eq.(3.2.10) for updating $\underline{c}_n$:

$$\underline{c}_{n+1} = \underline{c}_n + \frac{\underline{P}_n \underline{Q}^{-1} (\underline{I}-\underline{Q}) \underline{s}^*_{n+1} (x_{n+1}-\underline{s}^T_{n+1} \underline{c}_n)}{1 + \underline{s}^T_{n+1} \underline{P}_n \underline{Q}^{-1} (\underline{I}-\underline{Q}) \underline{s}^*_{n+1}}$$

Let $\underline{Q} = (\underline{I} + \alpha \underline{R}_{ss})^{-1}$ and $\underline{P}_n^{-1} = \underline{R}_{ss} + \underline{N}_u$

where $\underline{N}_u$ is the time varying component of $\underline{P}_n^{-1}$. For small α, we have argued in Section IV.1. that the power (or variance) of each component of $\underline{N}_u$ is "small".

$$\underline{P}_n \simeq \underline{R}_{ss}^{-1} - \underline{R}_{ss}^{-1} \underline{N}_u \underline{R}_{ss}^{-1} .$$

Let

$$ep_{n+1} = x_{n+1} - \underline{s}^T_{n+1} \hat{\underline{c}}_n .$$

Using these definitions and approxiamtions, the second term on the right hand side of eq.(3.2.10) can be replaced by

$$\alpha \underline{\varepsilon}_{n+1}^{*}\, ep_{n+1} - (\alpha \underline{R}_{ss}^{-1}\, \underline{N}_u\, ep_{n+1}\, \underline{s}^{*}_{n+1} +$$

$$\alpha^2 |\underline{s}_{n+1}|^2\, \epsilon\, p_{n+1} (\underline{I} - \underline{R}_{ss}^{-1}\, \underline{N}_u) s^{*}_{n+1}) /$$

$$(1 + \alpha\, \underline{s}_{n+1}^{T} (\underline{I} - \underline{R}_{ss}^{-1}\, \underline{N}_u) \underline{s}^{*}_{n+1})$$

Define the "misadjustment" vector

$$\underline{n}_{ma} = \frac{\alpha\, ep_{n+1} (\underline{R}_{ss}^{-1}\, \underline{N}_u + \alpha\, |\underline{s}_{n+1}|^2 (\underline{I} - \underline{R}_{ss}^{-1}\, \underline{N}_u)) \underline{s}^{*}_{n+1}}{(1 + \alpha |\underline{s}_{n+1}|^2 - \alpha\, \underline{s}_{n+1}^{T}\, \underline{R}_{ss}\, \underline{N}_u\, \underline{s}^{*}_{n+1}} \qquad (4.5.1)$$

Due to the fact that α is "small" and each element of $\underline{N}_u$ is "small", each element of $\underline{n}_{ma}$ is "small" compared to $\alpha\, ep_{n+1}\, s^{*}_{n+1}$. For special cases of $\{s_k\}$, such as pseudo random noise and sums of sinusoids, the expression for $\underline{n}_{ma}$ may be evaluated explicitly. For more general classes of input signals, it appears hopeless to study the properties of $\underline{n}_{ma}$ analytically.

With the reasonable assumption that $\underline{n}_{ma}$ is "small", we may ignore it altogether and write $\underline{c}_{n+1}$ (eq. 3.) as

$$\hat{\underline{c}}_{n+1} = \hat{\underline{c}}_n + \alpha\, \underline{s}^{*}_{n+1} (\underline{x}_{n+1} - \underline{s}_{n+1}^{T}\, \hat{\underline{c}}_n) \qquad (4.5.2)$$

As before, $\hat{\underline{c}}_o$ is arbitrary.

Eq. (4.5.2) will be called the "Approximate Least mean Squares" (ALMS) algorithm. We have derived this algorithm from the EWLS algorithm with $\underline{Q} = (\underline{I} + \alpha\, \underline{R}_{ss})^{-1}$ simply by omitting a

small "misadjustment" term. Except for this additional source of error, our analysis of the EWLS algorithm carries over to the ALMS algorithm also. Specifically, our discussions of the tracking behavior and the doppler sensitivity of the EWLS algorithm apply to the ALMS algorithm also.

The ALMS algorithm is similar to the so called "Least Mean Square" (LMS) algorithm developed by Widrow [51]. Widrow's version of the LMS algorithm reads

$$\hat{\underline{c}}_{n+1} = \hat{\underline{c}}_n + \underline{s}_n^* (\underline{x}_n - \underline{s}_n^T \hat{\underline{c}}_n) \qquad (4.5.3)$$

The discrepancy between (4.5.2) and (4.5.3) is probably due to Widrow's use of some dubious "approximations" such as "estimate the mean of the stochastic gradient by its instantaneous value" etc. (sic!) [51,54].

Several interesting signal processing applications of the LMS algorithm, such as prediction, antenna array design, noise cancellation and network modeling have been discussed by McCool [37] and Widrow and his coworkers [52,53]. Some other interesting applications of the LMS algorithm are in the areas of equalization in communication channels [20,40], echo cancellation [48] and, signal to noise ratio maximization [47,55], and instantaneous frequency estimation [23]. Analog equivalents of the LMS algorithms have also been considered by researchers [49].

Section V.

Conclusion

An exponentially weighted sequential least squares algorithm (EWLS), which is a modification of the classical least squares algorithm, has been developed. The Widrow-Hoff LMS algorithm is derived as an approximation to this algorithm. The only condition on the input signals $\{s_k\}$ is that the sequence $\{\underline{s}_k^* \ \underline{s}_k^T\}$ should have a finite positive definite mean and an integrable power spectrum. This implies that a wide variety of $\{\underline{s}_k\}$s, such as pseudorandom sequences, correlated noise like sequences and periodic pulsed FM signals are suitable for processing with the LMS adaptive algorithm. Analytic expressions (not all of which are of much practical value) have been derived to explain the significant properties of the EWLS (and hence, the LMS) algorithm, such as the "misadjustment noise", stability, error in tracking time varying parameters, effect of additive noise and the role of input signal. In particular, the usefulness of choosing an input signal sequence whose wideband ambiguity function is insensitive to doppler scaling becomes transparent.

Section VI.

Topics for further research.

1. The adaptive algorithms considered in this report have assumed that the observations are linear functions of the unknown parameters $\underline{c}_n$. The possibility of extending these results to nonlinear regression problems can be investigated. Section II.2. on Kalman filtering indicates how this problem can be approached for time invariant $\underline{c}$. The question is whether and how the nonlinear Kalman identification algorithms can be modified to apply to time varying situations. Possible applications include phase tracking in coherent communcation systems and simultaneous estimation of doppler seales and dispersion amplitudes in communications channels.

2. The estimation errors in adaptive filtering with the approximate least mean squares algorithm are highly dependent on α, the gain constant. The "misadjustment" noise and additive noise components of the error decrease, and the tracking error increases, with decreasing α. Presumably, there is an optimum choice of α for which the total error is minimized. In view of the complexity of the expressions for the three sources of error, it seems that a well chosen variable gain sequence is the only feasible way to find the optimum α.

3. We have shown that the "misadjustment" noise in the ALMS algorithm is related to the spectral density function of each term of $\{\underline{s}_k^* \underline{s}_k^T\}$. The misadjustment noise will be reduced if the sequences $(s_k^*(i)\ s_k(l))$ for each i and l have relatively

small frequency components in the low frequency region. It is useful to find signal sequences that have this property.

Section VI.

Topics for further research.

1. The adaptive algorithms considered in this report have assumed that the observations are linear functions of the unknown parameters $\underline{c}_n$. The possibility of extending these results to nonlinear regression problems can be investigated. Section II.2. on Kalman filtering indicates how this problem can be approached for time invariant $\underline{c}$. The question is whether and how the nonlinear Kalman identification algorithms can be modified to apply to time varying situations. Possible applications include phase tracking in coherent communcation systems and simultaneous estimation of doppler seales and dispersion amplitudes in communications channels.

2. The estimation errors in adaptive filtering with the approximate least mean squares algorithm are highly dependent on α, the gain constant. The "misadjustment" noise and additive noise components of the error decrease, and the tracking error increases, with decreasing α. Presumably, there is an optimum choice of α for which the total error is minimized. In view of the complexity of the expressions for the three sources of error, it seems that a well chosen variable gain sequence is the only feasible way to find the optimum α.

3. We have shown that the "misadjustment" noise in the ALMS algorithm is related to the spectral density function of each term of $\{\underline{s}_k^* \underline{s}_k^T\}$. The misadjustment noise will be reduced if the sequences $(s_k^*(i)\ s_k(l))$ for each i and l have relatively

References

1. Albert, A. E. and L. A. Gardner, Jr. "Stochastic Approximation and Nonlinear Regression." MIT Press. Cambridge, Mass. 1967.

2. Altes, R. A. and E. L. Titlebaum. "Bat Signals as Optimally Doppler Tolerant Waveforms." J. Acoust. Soc. Am. v-48, pp 1014-1020, 1970.

3. Astrom, K. J. and P. Eykhoff. "System identification - a survey." Automatica, v-7, pp 123-162, 1971.

4. Box, G. E. P. and G. M. Jenkins. "Some statistical aspects of adaptive optimization." J. Roy. Stat. Soc. Ser. B, v-24, pp 297-343, 1962.

5. Brennan, L. E. and I. S. Reed. "Theory of Adaptive Radar." IEEE Trans. v-AES-9, March, 1973.

6. Brown, J. E. "Adaptive Estimation in Nonstationary Environments." SEL-70-056 (Ph.D. Dissertation) Stanford Electronics Lab., Stanford Univ., Stanford, Calif. 1970.

7. Chien, Y.T. and K. S. Fu. "Learning in Nonstationary Environments using Dynamic Stochastic Approximation." Proc. 5th Allerton Conf. on Circuit and System Theory, pp 337-345, 1967.

8. _____ and _____. "On Bayesian Learning and Stochastic Approximation." IEEE Trans. v-SSC-3, pp 23-38, 1967.

9. Daniell, T. P. "Adaptive Estimation with Mutually Correlated Training samples." IEEE Trans. v-SSC-6, Jan. 1970.

10. Davis, H. T. and L. H. Koopmans. "Adaptive Prediction of stationary time series." Sankhya, ser. A. v-35, pt. 1, March, 1973.

11. Davisson, L. D. "A Theory of Adaptive Filtering." IEEE Trans. v-IT-12, April, 1966.

12. de Figureido, R. J. P. "Convergent Algorithms for Pattern Recognition in Nonlinearly Evolving Nonstationary Environments." Proc. IEEE (Letters) v-56, pp 188-189, 1968.

13. Fabian, V. "Stochastic Approximation of minima with improved speed." Ann. Math. Stat. v-38, pp 191-200, 1967.

14. _____. "Stochastic Approximation of constrained minima." Trans. 4th Prague Conference on Information Theory etc. pp 277-290, 1967.

15. Farden, D. C. "Stochastic Approximation with correlated data." Ph.D. Dissertation. Dept. of Elec. Engg., Colorado State Univ., Fort Collins, Colo. 1975.

16. Frost, O. L., "An algorithm for linearly constrained array processing." Proc. IEEE, v-60, Aug., 1972.

17. Fu, K. S., "Sequential methods in Pattern Recognition and Machine Learning." New York: Academic Press, 1968.

18. Gabriel, W. F. "Adaptive Arrays." Proc. IEEE, v-64, Feb., 1976.

19. Gardner, L. A. Jr. "Adaptive Predictors." Proc. 3rd Prague Conference on Information Theory, pp 123-192, 1964.

20. Gersho, A. "Adaptive Equalization of Highly Dispersive Channels for Data Transmission." Bell Syst. Tech. J. v-48, pp 55-70, 1969.

21. Glaser, E. M. "Signal Detection by Adaptive Filters." IRE Trans. v-IT-7, Apr. 1961.

22. Glover, J. R. "Adaptive Noise Cancellation of Sinusoidal Interference." Ph.D. Dissertation, Stanford Univ., Stanford, Calif. 1975.

23. Griffiths, L. J. "Rapid Measurement of Digital Instantaneous Frequency." IEEE Trnas. v-ASSP-23, pp 207-222, April, 1975.

24. _____. "A simple adaptive algorithm for real time processing of antenna arrays." Proc. IEEE, v-57, Oct. 1969.

25. Harris, F. J. "A maximum Entropy Filter." Rept. NUC TP 441, Naval Undersea Centre, San Diego, Calif. Jan. 1975.

26. Hoge, H. "Stochastic Algorithms in System Identification." Int. J. Control, v-17, pp 1121-1128, 1973.

27. Jones, R. H. "Exponential smoothing for multivariate time series." J. Roy. Stat. Soc., Ser. B, v-28, #1, 1966.

28. Kailath, T. "A view of three decades of linear filtering theory." IEEE Trans. v-IT-20, pp 146-181, March 1974.

29. _____. (Ed.) Special Issue on System identification and Time Series Analysis. IEEE Trans. v-AC-19, Dec. 1974.

30. Kalman, R. E. and R. S. Bucy. "New Results in Linear Filtering and Prediction Theory." Tnras. ASME, J. Basic. Eng. v-83, pp 95-108, 1961.

31. Kalman, R. E. "A new approach to linear filtering and predictive problems." J. Basic Eng. D, v-82, pp 35-45, 1960.

32. Kesten, H. "Accelerated Stochastic Approximation" Ann. Math. Stat. v-29, pp 41-59, 1958.

33. Lacoss, R. T. "Adaptive combining of wideband arrays for optimal reception." IEEE Trans. v-GE-6, pp 78-86, 1968.

34. Lee, R. C. K. "Optimal Estimation, Identification and Control." Cambridge, Mass. MIT Press, 1964.

35. Lukacs, E. "Stochastic Convergence." Lexington, Mass. D.C. Heath Co., 1968.

36. Makhoul, J. "Linear Prediction: A Tutorial Review Proc. IEEE, v-63, April 1975.

37. McCool, J. M. "Basic Principles of Adaptive Systems with Various Applications." NUC TN 837, Naval Undersea Centre, San Deigo, Calif., 1972.

38. Merriam, C. W. III "Automated Design of Control Systems." London: Gordon & Breach, 1974.

39. _____. "Adaptive Filters for modeling Undersea Communication Channels." Dept of Electrical Engr., Univ. of Rochester, Rochester, NY, 1975.

40. Proakis, J. G. "Equalization for Intersymbol Interference" in "Advances in Communications." v-IV (Ed.) A. V. Balakrishnan. New York, Academic Press, 1975.

41. Reed, I. S., J. D. Mallet and L. E. Brennan. "Rapid Convergence rate adaptive arrays." IEEE Trans. v-AES-10, Nov. 1974.

42. Reigler, R. L. and R. T. Compton. "Adaptive Array for Interference Rejection." Proc. IEEE, v-61, June, 1973.

43. Sage, A. P. and J. L. Melsa. "Estimation Theory with Applications to Communications and Control." New York, McGraw-Hill, 1971.

44. _____ and _____. "System Identification." New York, Academic Press, 1971.

45. Sakrison, D. J. "Stochastic Approximation" (in) "Advances in Communications," (Ed.) A. V. Balakrishnan, v-2, pp 51-102, 1966.

46. Scharf, L. L. "On Stochastic Approximation and the heirarchy of adaptive array algorithm" CDC Paper, #TA-26.

47. Schoenfeld, T. J. and M. Schwartz. "Rapidly Converging second order tracking algorithm." IEEE Trans. v-IT-17, pp 572-597, 1971.

48. Sondhi, M. M. "Adaptive Echo Canceller." Bell Syst. Tech. J., v-50, p 785, 1971.

49. Sondhi, M. M. and D. Mitra "New Rresults in the performance of a well known class of adaptive Filters." Proc. IEEE, Nov. 1976.

50. Wasan, N. T. "Stochastic Approximation." Cambridge Univ. Press, 1969.

51. Widrow, B. Adaptive Filters - I. Fundamentals." SEL-66-126; Stanford Electronics Lab., Stanford, Calif. 1966.

52. _____ et al. "Adaptive Antenna Systems." Proc. IEEE, v-12, pp 2143-2159, 1967.

53. _____ et al. "Adaptive Noise Cancellation - Principles and Applications." Proc. IEEE, v-63, pp 1692-1716, 1975.

54. _____ et al. "Stationary and Nonstationary Learning characteristics of the LMS adaptive filter." Proc. IEEE, v-64, Aug. 1976.

55. Winkler, L. and M. Schwartz. "Adaptive nonlinear optimization of the signal to noise ratio of an array subject to a constraint." J. Acoust. Soc. Am. v-52, 1972, pp 39-51.

56. Wilde, D. J. "Optimum Seeking Methods." Engelwood Cliffs, NJ: Prentice-Hall Inc., 1964.

57. Rao, C. R. "Linear Statistical Inference and its applications." New York: John Wiley, 1975.

58. Perl, J., D. Graupe and L. L. Scharf. "Convergent Identification Algorithms." Tech. Rept. #5, Dept. of Elec. Engr., Colorado State Univ., Fort Collins. Colo., March, 1974.