

1

Maximum Entropy Pole-Zero Estimation

Bruce R. Musicus

Allan M. Kabel

Technical Report 510

August 1985

Loan Copy Only

Massachusetts Institute of Technology
Research Laboratory of Electronics
Cambridge, Massachusetts 02139

Maximum Entropy Pole-Zero Estimation

Bruce R. Musicus

Allan M. Kabel

Technical Report 510

August 1985

Massachusetts Institute of Technology
Research Laboratory of Electronics
Cambridge, Massachusetts 02139

This work has been supported in part by the Advanced Research Projects Agency monitored by ONR under Contract N00014-81-K-0742 NR-049-506 and in part by the National Science Foundation under Grant ECS80-07102.

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION		1b. RESTRICTIVE MARKINGS	
2a. SECURITY CLASSIFICATION AUTHORITY		3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution unlimited	
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE			
4. PERFORMING ORGANIZATION REPORT NUMBER(S)		5. MONITORING ORGANIZATION REPORT NUMBER(S)	
6a. NAME OF PERFORMING ORGANIZATION Research Laboratory of Electronics Massachusetts Institute of Technology	6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION Office of Naval Research Mathematical and Information Scien. Div.	
6c. ADDRESS (City, State and ZIP Code) 77 Massachusetts Avenue Cambridge, MA 02139		7b. ADDRESS (City, State and ZIP Code) 800 North Quincy Street Arlington, Virginia 22217	
8a. NAME OF FUNDING/SPONSORING ORGANIZATION Advanced Research Projects Agency	8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER N00014-81-K-0742	
8c. ADDRESS (City, State and ZIP Code) 1400 Wilson Boulevard Arlington, Virginia 22217		10. SOURCE OF FUNDING NOS.	
11. TITLE (Include Security Classification) Maximum Entropy Pole-Zero Estimation		PROGRAM ELEMENT NO.	PROJECT NO.
			TASK NO. NR 049-506
			WORK UNIT NO.
12. PERSONAL AUTHOR(S) Bruce R. Musicus and Allan M. Kabel			
13a. TYPE OF REPORT Technical	13b. TIME COVERED FROM _____ TO _____	14. DATE OF REPORT (Yr., Mo., Day) August 1985	15. PAGE COUNT 94
16. SUPPLEMENTARY NOTATION Technical Report 510, Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, MA 02139			
17. COSATI CODES		18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)	
FIELD	GROUP	SUB. GR.	
19. ABSTRACT (Continue on reverse if necessary and identify by block number)			
<p>We describe a new Maximum Entropy pole-zero spectrum estimation method. The model is designed to achieve the maximum possible entropy subject to constraints on the first few correlation and cepstral values. The solution, which is in the form of an ARMA model, is based on solving a generalized, symmetric, almost-Toeplitz eigenvalue problem. We characterize the existence, uniqueness, stability and minimum phase properties of the solution, and categorize all possible occurrences of canceling pole-zero pairs. A search procedure based on a fast Levinson-like algorithm is given for estimating the model, and examples are presented to illustrate its performance. A special case of the method gives a model estimate similar to that of Pisarenko's harmonic retrieval problem.</p>			
20. DISTRIBUTION/AVAILABILITY OF ABSTRACT UNCLASSIFIED/UNLIMITED <input checked="" type="checkbox"/> SAME AS RPT. <input type="checkbox"/> OTIC USERS <input type="checkbox"/>		21. ABSTRACT SECURITY CLASSIFICATION Unclassified	
22a. NAME OF RESPONSIBLE INDIVIDUAL Kyra M. Hall RIE Contract Reports		22b. TELEPHONE NUMBER (Include Area Code) (617) 253-2569	22c. OFFICE SYMBOL

Maximum Entropy Pole-Zero Estimation

*Bruce R. Musicus*¹

*Allan M. Kabel*²

Room 36-797

Research Laboratory of Electronics
Massachusetts Institute of Technology

50 Vassar St.

Cambridge, Mass. 02139

Phone: (617) 253-8845

ABSTRACT

We describe a new Maximum Entropy pole-zero spectrum estimation method. The model is designed to achieve the maximum possible entropy subject to constraints on the first few correlation and cepstral values. The solution, which is in the form of an ARMA model, is based on solving a generalized, symmetric, almost-Toeplitz eigenvalue problem. We characterize the existence, uniqueness, stability and minimum phase properties of the solution, and categorize all possible occurrences of canceling pole-zero pairs. A search procedure based on a fast Levinson-like algorithm is given for estimating the model, and examples are presented to illustrate its performance. A special case of the method gives a model estimate similar to that of Pisarenko's harmonic retrieval problem.

August 30, 1985

Submitted to IEEE Transactions on Acoustics, Speech, Signal Proc.
Permission to publish this abstract separately is granted.

EDICS Category: 3.1.3.1 Parametric Spectral Analysis

¹This work has been supported in part by the Advanced Research Projects Agency monitored by ONR under Contract N00014-81-K-0742 NR-049-506 and in part by the National Science Foundation under Grant ECS80-07102.

²Now working at Sanders Associates, Nashua, New Hampshire

Maximum Entropy Pole-Zero Estimation

*Bruce R. Musicus*¹
*Allan M. Kabel*²

Room 36-797
Research Laboratory of Electronics
Massachusetts Institute of Technology
50 Vassar St.
Cambridge, Mass. 02139
Phone: (617) 253-8845

1. Introduction

Maximum Entropy has been suggested by numerous authors as a good objective measure for "optimally" modeling the power spectrum of a wide-sense stationary random process. In the original Maximum Entropy Spectral Analysis (MESA) formulation of Burg[1,2], the power spectrum $P(e^{j\omega})$ is chosen by maximizing the entropy function subject to constraints on several of the correlations of the model. Jaynes[3,4] has argued that the resulting maximum entropy model accurately describes the available information, but is maximally non-committal with regard to the unavailable information. In the case where the constraints are placed on a set of uniformly spaced correlations of a one-dimensional stationary process, Maximum Entropy analysis leads to an all-pole model whose coefficients may be found by a fast Levinson recursion algo-

¹This work has been supported in part by the Advanced Research Projects Agency monitored by ONR under Contract N00014-81-K-0742 NR-049-506 and in part by the National Science Foundation under Grant ECS80-07102.

²Now working at Sanders Associates, Nashua, New Hampshire

rithm. Modifications of the procedure, such as the covariance method, the forward/backward covariance method, and so forth, have been explored in depth by numerous authors[5, 6, 7].

In this paper, we consider a generalization of the Maximum Entropy Method of spectral estimation (MEM) in which we find the power spectrum with the largest entropy which matches both a set of correlation and a set of cepstral values. Lagunas-Hernandez *et. al.*[8] first showed that these constraints lead to an Autoregressive Moving-Average (ARMA) model for the power spectrum. Unable to solve for the pole and zero polynomial coefficients, however, they used an approximate solution technique having suboptimal performance. In this paper, we solve this problem exactly for the case of a one-dimensional process, with uniformly spaced correlation and cepstral lags centered about zero. We transform the problem into an equivalent generalized real symmetric almost-Toeplitz eigenvalue/eigenvector problem. The entropy of the model is related to the largest eigenvalue of this problem, and the pole polynomial is the corresponding eigenvector. The zero polynomial is then found by a simple recursion. The formulas are similar to the mixed first and second-order modeling procedure suggested by Mullis and Roberts[9]. Except for possible pole-zero cancellation, the pole polynomial is guaranteed to be stable. If the model we calculate is strictly minimum phase, then we show that it solves the constrained Maximum Entropy problem. Otherwise, we show that there does not exist any finite, strictly positive power spectrum which exactly maximizes the entropy. We speculate in this case, however, that our model is the weak limit of a sequence of minimum phase models which match

all given constraints, and which asymptotically achieve the maximum possible entropy.

Because the eigenvector problem has an almost-Toeplitz structure with displacement rank 3, there exists a fast Levinson-like algorithm for finding the eigenvector once the eigenvalue is known. This same algorithm can be used iteratively to help search for the maximum eigenvalue. A side benefit of this approach is that, at least in theory, we can detect when the model has canceling pole-zero pairs caused by choosing an excessively high model order.

When the number of known cepstra is equal to or larger than the number of known correlations, and all known cepstra have value zero, then we show that our MEM model yields the same answer as Pisarenko's harmonic retrieval algorithm[10] We conclude with several examples illustrating the performance of this pole-zero maximum entropy estimator on a variety of simulated data.

2. Derivation of an ARMA Model

Suppose we observe a segment of N data samples $x[0], \dots, x[N-1]$ drawn from a zero-mean, stationary, ergodic, complex-valued Gaussian random process with unknown power spectrum $P(z)$. (Though we only consider complex-valued data, the development for real-valued data would be virtually identical.) The correlations $R[n]$ of this process are defined by:

$$R[n] = E \left[x^*[k]x[k+n] \right] \quad \text{for any } k \quad (2.1)$$

$$= \int_{-\pi}^{\pi} P(e^{j\omega}) e^{j\omega n} \frac{d\omega}{2\pi} \quad (2.2)$$

The power spectrum $P(z)$ of this process is the z-transform of the correlations:

$$P(z) = \sum_{n=-\infty}^{\infty} R[n]z^{-n} \quad (2.3)$$

We define the cepstrum of the process $x[n]$ as:

$$c[n] = \int_{-\pi}^{\pi} \log P(e^{j\omega}) e^{j\omega n} \frac{d\omega}{2\pi} \quad (2.4)$$

Using the available time series data, we would like to estimate the power spectrum $P(z)$. The classical Blackman-Tukey[11] approach to power spectrum estimation first estimates the correlations from the data by a formula such as:

$$\hat{R}[n] = \frac{1}{N} \sum_{k=0}^{N-n-1} x^*[k]x[k+n] \quad (2.5)$$

Unfortunately, given only the available segment of N data points, it will not be possible to directly estimate the correlations beyond lag $N-1$. Furthermore, for n close to N , only $N - |n|$ terms are available in the sum for estimating $R[n]$, and so these estimates will have large variance. The classical power spectrum estimation approach multiplies the estimated correlations by a tapered window which attenuates the high order, unreliable lags. Applying a Fourier Transform then gives a smoothed power spectral estimate. Increasing the lag window length improves the spectral resolution, but also increases the variance of the spectrum.

Burg[1,2] suggested an alternative procedure for power spectrum estimation which often achieves higher resolution with less variance than the classical estimates. Assume

that our estimates of the first $p + 1$ low order correlations $\hat{R}[0], \dots, \hat{R}[p]$ are exactly correct, but assume that the remaining correlations are unknown. In general, there will be an infinite number of power spectra which match the known correlations. Burg suggested that a reasonable estimation approach would be to choose the power spectrum which matches the known correlations, but otherwise has maximum entropy H . For a stationary Gaussian random process, the average entropy can be shown to be proportional to:

$$H = \int_{-\pi}^{\pi} \log P(e^{j\omega}) \frac{d\omega}{2\pi} \quad (2.6)$$

Maximizing this entropy formula subject to the correlation constraints yields an Autoregressive (AR) model for the power spectrum. When the data consists of uniformly sampled correlations of a one-dimensional process, this model is easily calculated by a fast Levinson recursion algorithm which is guaranteed to yield a stable all-pole polynomial estimate. Numerous studies have proven the high resolution capabilities of this spectrum modeling approach, and have shown that the spectral estimate is a good model of the envelope of the actual power spectrum[6].

In this paper we consider a straightforward extension to Burg's Maximum Entropy procedure which leads to an Autoregressive Moving Average (ARMA) model. Suppose that we know the exact values of the first $p + 1$ correlations of the process, $R[0], \dots, R[p]$. Also suppose that we know the exact values of the first q values of the cepstrum of the process, $c[1], \dots, c[q]$. (Note that the entropy is equal to the zeroth cepstral coefficient $c[0]$; this is why we assume that we know only the values

$c[1], \dots, c[q]$.) The additional cepstral information should be useful in generating a much higher quality spectral estimate than we could achieve with just correlation data alone. For example, in applications such as the analysis of voiced segments of speech, we would like good estimates of the envelope of the signal power spectrum, without the fine structure introduced by the voiced excitation. It is well known that the low order cepstral coefficients accurately reflect this envelope information without much degradation from the impulse train excitation. This cepstral property has been extensively exploited in homomorphic systems such as those introduced by Kopec, Oppenheim and Tribolet[12], and Yegnanarayana[13].

We now calculate the power spectrum with the largest entropy (2.6) but which satisfies the constraints (2.2) and (2.4). Recognize that $R[-n] = R^*[n]$ and $c[-n] = c^*[n]$. Introduce Lagrange multipliers $\{\theta_k\}_{k=-p}^p$ and $\{\mu_k\}_{k=-q}^q$, and form the Lagrangian L by adding multiples of the constraints (2.2) and (2.4) to the entropy H :

$$L = \mu_0 \int_{-\pi}^{\pi} \log P(e^{j\omega}) \frac{d\omega}{2\pi} + \sum_{k=-p}^p \theta_k \left\{ \int_{-\pi}^{\pi} P(e^{j\omega}) e^{j\omega k} \frac{d\omega}{2\pi} - R[k] \right\} \quad (2.7)$$

$$+ \sum_{|k|=1}^q \mu_k \left\{ \int_{-\pi}^{\pi} \log P(e^{j\omega}) e^{j\omega k} \frac{d\omega}{2\pi} - c[k] \right\}$$

Suppose that a solution to the Maximum Entropy problem exists, $\hat{P}(e^{j\omega})$, which exactly matches the correlations and the cepstra, and which is finite and strictly positive. Then by the Euler-Lagrange Multiplier Theorem[14] this solution must be a

critical point of the Lagrangian with respect to variations in $P(e^{j\omega})$ and variations in the multipliers θ_k and μ_k . Using variational calculus, therefore, a necessary condition which the Maximum Entropy solution must satisfy is:

$$0 = \frac{\mu_0}{P(e^{j\omega})} + \sum_{|k|=0}^p \theta_k e^{j\omega k} + \sum_{|k|=1}^q \mu_k \frac{1}{P(e^{j\omega})} e^{j\omega k} \quad (2.8)$$

(Loosely speaking, we differentiate L with respect to each point of the function $P(e^{j\omega})$, and set the derivative to zero.) Solving (2.8) shows that if a strictly positive solution exists to the Maximum Entropy problem, $\hat{P}_{p,q}(e^{j\omega}) > 0$ for all ω , then it must have the form of a rational polynomial, or in other words, an Autoregressive Moving Average (ARMA) model:

$$\hat{P}_{p,q}(z) = \frac{\sum_{k=-q}^q \mu_k z^k}{\sum_{k=-p}^p \theta_k z^k} \quad (2.9)$$

The Lagrange multiplier coefficients $\{\theta_k\}$ and $\{\mu_k\}$ must be chosen so that the constraints (2.2) and (2.4) are satisfied. It is well known that linear time invariant systems satisfying finite order difference equations are naturally described by ARMA models [15]; thus this approach to spectral estimation has the potential of yielding accurate models for many physical systems.

3. Determining the ARMA Model's Coefficients

This MEM modeling technique was first published by Lagunas-Hernandez, *et al*[8]. In this original paper, however, only approximate solutions for the model parameters $\{\theta_k, \mu_k\}$ were suggested. We have found, however, that it is possible to

transform this problem into the form of a generalized eigenvalue/eigenvector problem which is comparatively easy to solve.

Assume that a solution $P(e^{j\omega})$ exists to our constrained Maximum Entropy problem which satisfies the Paley-Wiener condition for discrete signals:[16]

$$\int_{-\pi}^{\pi} \left| \log P(e^{j\omega}) \right| \frac{d\omega}{2\pi} < \infty \quad (3.1)$$

Then we can always factor the power spectrum into a product of a gain γ^2 times a minimum phase factor $G(z)$ times a maximum phase factor $G^*(1/z^*)$:

$$P(z) = \gamma^2 G(z) G^*(1/z^*) \quad (3.2)$$

where $G(z)$ is causal and stable, with leading coefficient of 1, and with a causal and stable inverse.

$$G(z) = 1 + \sum_{n=1}^{\infty} g[n] z^{-n} \quad (3.3)$$

We define $g[0] \equiv 1$, and $g[n] \equiv 0$ for $n < 0$. Because the MEM power spectrum (2.9) must be a finite order rational polynomial, $G(z)$ can be factored into a ratio of two minimum phase polynomials:

$$G(z) = \frac{B_q(z)}{A_p(z)} \quad (3.4)$$

where:

$$A_p(z) = 1 + \sum_{n=1}^p a_{n,p} z^{-n} \quad (3.5)$$

$$B_q(z) = 1 + \sum_{n=1}^q b_{n,q} z^{-n} \quad (3.6)$$

All the roots of $A_p(z)$ and $B_q(z)$ must be inside the unit circle. It is convenient to define $a_{0,p} \equiv 1$ and $b_{0,q} \equiv 1$.

Oppenheim and Schafer[17] derived a simple formula for minimum phase models which relates the cepstral coefficients, $c[n]$, to the coefficients of the impulse response, $g[n]$. To derive this relationship, take the log of (3.2):

$$\sum_{n=-\infty}^{\infty} c[n]z^{-n} = \log \gamma^2 + \log G(z) + \log G^*(1/z^*) \quad (3.7)$$

Since $G(z)$ is minimum phase, the polynomial $\log G(z)$ is right sided with non-zero coefficients starting at lag 1. Similarly, $\log G^*(1/z^*)$ is a left sided polynomial starting at lag -1. Matching terms with equal powers of z on both sides of (3.7) then gives:

$$\log G(z) = \sum_{n=1}^{\infty} c[n]z^{-n} \quad (3.8)$$

and

$$c[0] = H = \log \gamma^2 \quad (3.9)$$

Equation (3.8) specifies a mapping between the coefficients of the minimum phase polynomial $G(z)$ and the cepstral coefficients $c[n]$. Differentiating both sides with respect to z^{-1} gives:

$$\sum_{n=1}^{\infty} n c[n]z^{-n+1} = \frac{1}{G(z)} \sum_{n=1}^{\infty} n g[n]z^{-n+1} \quad (3.10)$$

Multiplying (3.10) by $G(z)$ and equating terms with equal powers of z on both sides of the equation thus gives a recursive formula for the coefficients $g[n]$ in terms of the cepstral coefficients $c[n]$:

$$g[0] = 1 \quad (3.11)$$

$$g[n] = \frac{1}{n} \sum_{k=1}^n k c[k] g[n-k] \quad \text{for } n = 1, 2, \dots$$

Thus knowing the cepstral coefficients $c[1], \dots, c[q]$ is exactly equivalent to knowing the first $q+1$ coefficients of the minimum phase factor, $g[0], \dots, g[q]$ (where $g[0]=1$). Furthermore, the entropy H , which is equal to the zeroth cepstral coefficient $c[0]$, is just the log of the model gain γ^2 . We may conclude that maximizing entropy subject to knowledge of the correlations $R[0], \dots, R[p]$ and cepstra $c[1], \dots, c[q]$ is thus equivalent to finding a minimum phase pole-zero model with p poles and q zeroes, which has the largest possible gain γ^2 , and which also matches the correlations $R[0], \dots, R[p]$ and impulse response coefficients $g[0], \dots, g[q]$.

Now to find the pole and zero coefficients. Multiplying both sides of (3.4) by $A_p(z)$ gives:

$$A_p(z)G(z) = B_q(z) \quad (3.12)$$

Equating like powers of z on both sides, and recognizing that $g[0], \dots, g[q]$ are known, gives $q+1$ linear equations relating the $b_{n,q}$ and $a_{n,p}$ coefficients.

$$\begin{pmatrix} b_{q,q} \\ \vdots \\ b_{1,q} \\ 1 \end{pmatrix} = \begin{pmatrix} g[q-p] & \cdot & g[q] \\ & \cdot & \vdots \\ & g[0] & g[1] \\ 0 & & g[0] \end{pmatrix} \begin{pmatrix} a_{p,p} \\ \vdots \\ a_{1,p} \\ 1 \end{pmatrix} \quad (3.13)$$

or:

$$\mathbf{b} = \mathbf{G}_{p,q} \mathbf{a} \quad (3.14)$$

where we define the vectors \underline{b} , \underline{a} and the matrix $G_{p,q}$ in the obvious way. Note that $G_{p,q}$ is a $(q+1) \times (p+1)$ rectangular upper triangular Toeplitz matrix. Now multiply both sides of (3.2) by $A_p(z)$ and substitute (3.4) to get:

$$A_p(z)P(z) = \gamma^2 B_q(z)G^*(1/z^*) \quad (3.15)$$

Since $P(z) = \sum_{n=-\infty}^{\infty} R[n]z^{-n}$, and we know $R[0], \dots, R[p]$ and $g[0], \dots, g[q]$, by

equating terms with like powers of z on both sides of (3.15), we can construct $p+1$ more linear equations relating the $a_{n,p}$ and $b_{n,q}$ coefficients:

$$\begin{pmatrix} R[0] & \cdot & R[p] \\ \vdots & \cdot & \cdot \\ R[-p] & \cdot & R[0] \end{pmatrix} \begin{pmatrix} a_{p,p} \\ \vdots \\ a_{1,p} \\ 1 \end{pmatrix} = \gamma^2 \begin{pmatrix} g^*[q-p] & \cdot & 0 \\ \vdots & \cdot & \cdot \\ g^*[q] & \cdot & g^*[0] \end{pmatrix} \begin{pmatrix} b_{q,q} \\ \vdots \\ b_{1,q} \\ 1 \end{pmatrix} \quad (3.16)$$

or:

$$R_p \underline{a} = \gamma^2 G_{p,q}^H \underline{b} \quad (3.17)$$

where we define R_p as the $(p+1) \times (p+1)$ Toeplitz correlation matrix on the left of (3.16), and where $G_{p,q}^H = G_{p,q}^{T^*}$ is the Hermitian (complex conjugate transpose) of $G_{p,q}$. We can eliminate the zero polynomial coefficients \underline{b} by substituting (3.13), giving:

$$K_\gamma(p,q) \underline{a} = \underline{0} \quad (3.18)$$

$$\text{where } K_\gamma(p,q) = R_p - \gamma^2 G_{p,q}^H G_{p,q}$$

Computing the solution to (3.18) requires finding a value of $\hat{\gamma}^2$ such that $K_{\hat{\gamma}}(p,q)$ has a non-trivial null space. Then $\hat{\underline{a}}$ is an appropriately scaled member of this null space.

This matrix $K_{\gamma}(p, q)$ has a very interesting structure. Suppose that there exists some minimum phase polynomial $\bar{G}(z)$ and gain $\bar{\gamma}^2$ such that this polynomial matches the given leading impulse response and correlation values:

$$\bar{g}[n] = g[n] \quad \text{for } n=0, \dots, q \quad (3.19)$$

$$\bar{\gamma}^2 \sum_{k=0}^{\infty} \bar{g}^*[k] \bar{g}[k+n] = R[n] \quad \text{for } n=0, \dots, p$$

We do *not* assume that the polynomial $\bar{G}(z)$ has maximum entropy; in general, there will be many such polynomials. Substituting into (3.18) shows that the $K_{\gamma}(p, q)$ matrix must equal:

$$K_{\bar{\gamma}}(p, q) = \bar{\gamma}^2 \sum_{k=q+1}^{\infty} \begin{pmatrix} \bar{g}^*[k-p] \\ \vdots \\ \bar{g}^*[k] \end{pmatrix} \left(\bar{g}[k-p] \cdots \bar{g}[k] \right) \quad (3.20)$$

The matrix $K_{\bar{\gamma}}(p, q)$ is thus equal to the covariance matrix formed from the tail of the $\bar{g}[n]$. This matrix $K_{\bar{\gamma}}(p, q)$ must therefore be symmetric and positive semi-definite.

We conclude that the MEM gain $\hat{\gamma}^2$ must be chosen so that $K_{\gamma}(p, q)$ is positive semi-definite with a non-trivial null space, because otherwise the MEM model impulse response $\hat{g}[n]$ could not possibly meet the constraints.

To calculate the MEM gain, we multiply (3.18) through by $1/\gamma^2$, thus converting the equation into the following generalized, conjugate symmetric eigenvalue problem:

Solve:

$$G_{p,q}^H G_{p,q} x_i = \lambda_i R_p x_i \quad (3.21)$$

for all generalized eigenvalues λ_i and corresponding eigenvectors \mathbf{x}_i . (Note that the conventional form of the generalized real symmetric eigenvalue problem requires the matrix on the right hand side, \mathbf{R}_p , to be conjugate symmetric and positive definite [18, 19]). The MEM gain should then be set to the inverse of one of these generalized eigenvalues, and the pole coefficient vector should be set to the corresponding eigenvector. But which eigenvalue do we choose for $\hat{\gamma}^2$? Appendix A proves:

Lemma 1 Assume that $\mathbf{R}_p > 0$. Then the generalized eigenvalue problem (3.21) has $p + 1$ non-negative eigenvalue solutions λ_i . Let λ_{\max} be the largest eigenvalue; this will always be strictly positive, $\lambda_{\max} > 0$. Then for γ^2 in the range $0 \leq \gamma^2 < 1/\lambda_{\max}$, the matrix $\mathbf{K}_\gamma(p, q)$ will be strictly positive-definite. For $\gamma^2 = 1/\lambda_{\max}$ the matrix $\mathbf{K}_\gamma(p, q)$ is positive semi-definite with a non-trivial null space. For $\gamma^2 > 1/\lambda_{\max}$, the matrix $\mathbf{K}_\gamma(p, q)$ is not positive semi-definite.

Using this result, Appendix A proves:

Theorem 1 Assume that $\mathbf{R}_p > 0$. Then:

- A) $\hat{\gamma}^2 = 1/\lambda_{\max}$ is the only choice for the gain which satisfies the requirements of the Maximum Entropy problem.
- B) Let $P(z) = \gamma^2 G(z)G^*(1/z^*)$ be any power spectrum with minimum phase factor $G(z)$, which matches the leading impulse response coefficients $g[0], \dots, g[q]$ and the correlations $R[0], \dots, R[p]$. Then the entropy of this model, which has value $H = \log \gamma^2$, is bounded above by $H \leq \log \hat{\gamma}^2$.

We will call the model with gain $\hat{\gamma}^2 = 1/\lambda_{\max}$, and pole/zero polynomials $\hat{A}_p(z)$, $\hat{B}_q(z)$ which satisfy (3.18) and (3.13) the *MEM* (p, q) solution. We can calculate this *MEM* (p, q) power spectrum by the following procedure:

- 1) Recursively compute the minimum phase coefficients $g[0], \dots, g[q]$ from the cepstral coefficients $c[1], \dots, c[q]$ using (3.11).

- 2) Construct the matrices $G_{p,q}$ and R_p with $(n,m)^{th}$ elements $[G_{p,q}]_{n,m} = g[q-p+m-n]$ and $[R_p]_{n,m} = R[m-n]$.
- 3) Solve for the largest generalized eigenvalue λ_{\max} solution to (3.21) and its corresponding eigenvector x_{\max} . Set $\hat{\gamma}^2 = 1/\lambda_{\max}$. Set \hat{a} equal to a corresponding eigenvector, scaled so that $\hat{a}_{0,p} = 1$.
- 4) Recursively compute the \hat{b} coefficients from (3.13).
- 5) The *MEM* (p,q) power spectrum then has the form

$$\hat{P}_{p,q}(z) = \hat{\gamma}^2 \frac{\hat{B}_q(z) \hat{B}_q^*(1/z^*)}{\hat{A}_p(z) \hat{A}_p^*(1/z^*)} \quad (3.22)$$

The special cases of pure Autoregressive (AR) or pure Moving Average (MA) models are particularly easy to solve. For the *MEM* ($p,0$) all-pole modeling problem:

$$K_{\gamma}(p,0) = R_p - \gamma^2 \begin{pmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{pmatrix} \begin{pmatrix} 0 & \cdots & 0 & 1 \end{pmatrix} \quad (3.23)$$

and equation (3.18) reduces to solving:

$$R_p \begin{pmatrix} a_{p,p} \\ \vdots \\ a_{1,p} \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ \gamma^2 \end{pmatrix} \quad (3.24)$$

This is exactly the problem Burg suggested, giving a p pole AR model matching the first $p+1$ correlations $R[0], \dots, R[p]$. Because R_p is Toeplitz, this may be solved with a fast Levinson algorithm[5].

Another special case that is easy to solve is the *MEM* $(0,q)$ all-zero modeling problem, when only the zeroth correlation coefficient $R[0]$ is known, together with q cepstral coefficients $c[1], \dots, c[q]$. Equation (3.18) gives:

$$\hat{\gamma}^2 = \frac{R[0]}{\sum_{n=0}^q |g[n]|^2} \quad (3.25)$$

and the optimal $\hat{B}(z)$ polynomial is equal to the leading impulse response coefficients:

$$\hat{b}_{n,q} = g[n] \quad \text{for } n=0, \dots, q \quad (3.26)$$

The *MEM* (p,q) solution is very similar to the least squares, mixed first and second order information problem suggested by Mullis and Roberts for filter design[9]. The chief difference is that in their problem, the gain γ is effectively assumed to be known, and the model is not forced to match the given correlations. As pointed out by Mullis and Roberts, this method is also quite similar to Prony's ARMA modeling method, in which we choose the pole polynomial which optimally linearly predicts the semi-infinite tail of the impulse response data, and then find the zeroes from the leading impulse response coefficients. These similarities allow us to carry over many of the stability results and fast algorithms that have been developed for these other methods.

4. Properties of the *MEM* (p,q) solution

The matrix $K_\gamma(p,q)$ has an interesting structure which allows us to prove a variety of interesting properties for this MEM algorithm. From (3.18) we see that the $(n,m)^{th}$ element in this matrix has the formula:

$$\left[K_{\gamma}(p, q) \right]_{n, m} = R[m-n] - \gamma^2 \sum_{k=0}^q g^*[k-p+n]g[k-p+m] \quad (4.1)$$

where we define $g[n] = 0$ for $n < 0$. Careful examination of this formula shows that we can recursively define $K_{\gamma}(p, q)$ in terms of the lower order matrix $K_{\gamma}(p-1, q-1)$ in at least two different ways:

$$K_{\gamma}(p, q) = \begin{pmatrix} K_{\gamma}(p-1, q-1) & * \\ * & * \end{pmatrix} \quad (4.2)$$

and

$$K_{\gamma}(p, q) = \begin{pmatrix} \bar{R}[0] & \bar{R}[1] & \cdots & \bar{R}[p] \\ \bar{R}[-1] & & & \\ \vdots & K_{\gamma}(p-1, q-1) & & \\ \bar{R}[-p] & & & \end{pmatrix} - \gamma^2 \begin{pmatrix} g^*[q-p] \\ \vdots \\ g^*[q] \end{pmatrix} \begin{pmatrix} g[q-p] & \cdots & g[q] \end{pmatrix} \quad (4.3)$$

$$\text{where } \bar{R}[n] = \begin{cases} R[n] & \text{if } q \leq p \\ R[n] - \gamma^2 \sum_{k=0}^{q-p-1} g^*[k]g[k+n] & \text{if } q > p \text{ and } n \geq 0 \\ \bar{R}^*[-n] & \text{if } n < 0 \end{cases}$$

where $*$ denotes elements whose values are not of immediate interest, and where we define $K_{\gamma}(n, m) \equiv R_n$ if $m < 0$. These relationships indicate that $K_{\gamma}(p, q)$ is an almost-Toeplitz matrix with displacement rank of three[20,21]. It is this structure which Mullis and Roberts exploited to derive a fast Levinson-like algorithm for finding the vector \hat{a} once $\hat{\gamma}^2$ is known. This structure also allows us to analyze the existence, uniqueness, and stability of the *MEM* (p, q) solution. Some of the theorems that follow are improved versions of theorems in [9]; others are new. For clarity, we first

state the results for the special (and most common case) where the multiplicity, r , of the largest eigenvalue λ_{\max} is $r = 1$. Appendix B proves the following:

Theorem 2A Assume that $R_p > 0$, and that the maximum eigenvalue solution λ_{\max} to (3.21) has multiplicity $r = 1$. Let $\hat{\gamma}^2 = 1/\lambda_{\max}$. Then:

- A) $K_{\hat{\gamma}}(p, q) \geq 0$, and its null space has dimension 1. For $1 \leq s \leq p$, $K_{\hat{\gamma}}(p-s, q-s) > 0$.
- B) The *MEM* (p, q) problem has a unique solution $\hat{\gamma}$, $\hat{A}_p(z)$, $\hat{B}_q(z)$.

Appendix C proves:

Theorem 3 Under the same assumptions as in Theorem 2:

- A) The *MEM* (p, q) model pole polynomial $\hat{A}_p(z)$ is stable, with all poles inside or on the unit circle.
- B) If any poles are actually on the unit circle, then they will be canceled by matching zeroes in the zero polynomial $\hat{B}_q(z)$.
- C) Except for possible canceling pole-zero pairs on the unit circle, the *MEM* (p, q) model will have no other canceling pole-zero pairs.

Unfortunately, it is quite easy to construct cases in which the *MEM* (p, q) solution places canceling pole-zero pairs on the unit circle. Consider, for example, the following *MEM* (1,1) problem: $R[0] = 2$, $R[1] = 1$, $c[1] = 0$. Using the recursive equation (3.11), we derive $g[0] = 1$, and $g[1] = 0$. Constructing the R_1 and $G_{1,1}$ matrices and solving the generalized eigenvalue problem (3.21), we find that $\lambda_0 = 1$ and $\lambda_1 = 1/3$.

Thus $r = 1$, and the unique *MEM* (1,1) solution is:

$$\begin{aligned} \hat{\gamma}^2 &= 1/\lambda_{\max} = 1 & (4.4) \\ \hat{A}_1(z) &= 1 - z^{-1} \\ \hat{B}_1(z) &= 1 - z^{-1} \end{aligned}$$

This model has a canceling pole-zero pair at 1. Its leading impulse response coefficients are $\hat{g}[0] = 1$ and $\hat{g}[1] = 0$. Its power spectrum is $\hat{P}_{1,1}(z) = 1$, and its correlations are $\hat{R}[0] = 1$ and $\hat{R}[1] = 0$. Note that although the impulse response values are correct, the correlation values do not match the given values $R[0]$ and $R[1]$. We will return to this problem in theorem 5.

When the multiplicity r of the largest eigenvalue λ_{\max} in (3.21) is greater than 1, then the situation is more complicated. In this case, it turns out that it is the $MEM(p-r+1, q-r+1)$ problem which has a unique solution. The $MEM(p, q)$ solution differs from the $MEM(p-r+1, q-r+1)$ solution only in that it may contain up to $r-1$ extra canceling pole-zero pairs which may be placed anywhere. In more detail, Appendix B proves:

Theorem 2B Assume that $R_p > 0$, and that the maximum eigenvalue solution λ_{\max} to (3.21) has multiplicity r . Let $\hat{\gamma}^2 = 1/\lambda_{\max}$. Then:

- A) $r \leq \min(p, q) + 1$
- B) For $0 \leq s < r$, $K_{\hat{\gamma}}(p-s, q-s) \geq 0$, and the null spaces of these matrices have dimension $r-s$. For $r \leq s \leq p$, $K_{\hat{\gamma}}(p-s, q-s) > 0$.
- C) The $MEM(p-r+1, q-r+1)$ problem has a unique solution $\hat{\gamma}$, $\hat{A}_{p-r+1}(z)$, $\hat{B}_{q-r+1}(z)$.
- D) For s in the range $0 \leq s < r$, all solutions of the $MEM(p-s, q-s)$ problem have the same gain $\hat{\gamma}$ as the $MEM(p-r+1, q-r+1)$ solution.
- E) For s in the range $0 \leq s < r$, the pole and zero polynomials $\hat{A}_{p-s}(z)$ and $\hat{B}_{q-s}(z)$ are solutions to the $MEM(p-s, q-s)$ problem if and only if $\hat{A}_{p-s}(z) = \hat{A}_{p-r+1}(z)\phi_{r-s-1}(z)$ and $\hat{B}_{q-s}(z) = \hat{B}_{q-r+1}(z)\phi_{r-s-1}(z)$, where $\phi_{r-s-1}(z)$ is a polynomial of order up to $r-s-1$ with leading coefficient $\phi_0 = 1$. In other words, the $MEM(p-s, q-s)$ solution is unique except for having up to $r-s-1$ canceling pole-zero pairs, $\phi_{r-s-1}(z)$.
- F) For s in the range $0 \leq s < r$, the $MEM(p-s, q-s)$ power spectrum $\hat{P}_{p-s, q-s}(z)$ is uniquely defined, and is equal to the $MEM(p-r+1, q-r+1)$ power spectrum $\hat{P}_{p-r+1, q-r+1}(z)$.

Theorem 3 applies to the $MEM(p-r+1, q-r+1)$ model. Because the $r-1$ canceling pole-zero pairs in the $MEM(p, q)$ solution may be located arbitrarily, however, it is impossible to guarantee that the $MEM(p, q)$ model has all its poles inside the unit circle.

Does the $MEM(p, q)$ solution match the given constraints and solve the original Maximum Entropy problem? The answer is "yes" if the solution is strictly minimum phase, but "no" otherwise. In more detail, Appendix D proves (for any multiplicity r):

Theorem 4 Assume that $R_p > 0$ and that λ_{\max} has multiplicity r . Let $\hat{\gamma}^2 = 1/\lambda_{\max}$, and let $\hat{A}_p(z)$, $\hat{B}_q(z)$ be any pole-zero solution to the $MEM(p, q)$ problem. Then:

- A) The first $q+1$ coefficients of the causal impulse response of $\hat{B}_q(z)/\hat{A}_p(z)$ are equal to $g[0], \dots, g[q]$.
- B) If the poles of $\hat{A}_p(z)$ are strictly inside the unit circle, then the first $p+1$ correlations of the model are equal to $R[0], \dots, R[p]$.
- C) If the zeroes and the poles are all strictly inside the unit circle, so that the model is strictly minimum phase, then the first q cepstral coefficients of the model are equal to $c[1], \dots, c[q]$. Furthermore, the model is the solution to our original constrained Maximum Entropy problem, with entropy equal to $H = \log \hat{\gamma}^2$.

Appendix D qualifies these results with the following:

Theorem 5 Under the same assumptions as in theorem 4:

- A) If the model has poles on the unit circle at $\rho_i = e^{j\omega_i}$ for some set of ω_i , then these will be canceled by matching zeroes at ρ_i . The model correlations $\hat{R}[n]$ will still be well defined, but they may no longer match the given correlations, and the error may have the form:

$$R[n] - \hat{R}[n] = \sum_i \sum_{j=0}^{s_i-1} \mu_{i,j} n^{2j} \rho_i^n \quad (4.5)$$

where the $\mu_{i,j}$ are some set of weights, and where s_i is the multiplicity of root ρ_i .

- B) If the poles are strictly inside the unit circle, but one or more zeroes are strictly outside the unit circle, then the model will not match the cepstral values $c[1], \dots, c[q]$, and the model entropy will be larger than $\log \hat{\gamma}^2$.
- C) For either cases A) or B), there does not exist a finite and strictly positive power spectrum meeting the Paley-Wiener condition which matches the given correlation and cepstral values, and which also achieves the maximum possible entropy.

Note that in our $MEM(1,1)$ example above, with the canceling pole-zero pair at $\rho_0 = 1$, the correlation matching error has exactly the form predicted by this theorem, $R[n] - \hat{R}[n] = 1^n$.

To summarize, if the solution to the $MEM(p, q)$ problem is strictly minimum phase, then it is the unique power spectrum solution to the constrained maximum entropy problem. On the other hand, there are a variety of situations in which we may encounter difficulties:

- 1) If the correlation matrix R_p is only positive semi-definite, then these theorems do not apply. In this case, only one power spectrum exists which matches the known correlations, and this is a line spectrum formed from p impulses[10]. The entropy and the cepstral coefficients of this spectrum are not well defined.
 - 2) If the maximum generalized eigenvalue solution λ_{\max} to (3.21) has multiplicity $r > 1$, then canceling pole-zero pairs may occur outside the unit circle, and so $\hat{A}_p(z)$ may be unstable.
 - 3) The solution to the $MEM(p - r + 1, q - r + 1)$ problem, though unique, could have one or more canceling pole-zero pairs located on the unit circle. In this case, the model may not match the given correlations or cepstra.
-

- 4) Even if all the poles of the $MEM(p-r+1, q-r+1)$ solution are strictly inside the unit circle, one or more zeroes may be outside the unit circle. In this case, the model will match the correlations $R[0], \dots, R[p]$ and the causal impulse response will match $g[0], \dots, g[q]$, but since the model is not minimum phase, it will not match the cepstra $c[1], \dots, c[q]$.

In practice, when we estimate the correlations and cepstra from finite segments of data, this fourth case occurs quite often. As a result, in practice the $MEM(p, q)$ solution often does not solve the original Maximum Entropy problem. We discuss this further in section 7, where we present several as yet unproven speculations. In practice, when excessively high model orders are used, we should also note that it is common to see nearly canceling pole-zero pairs appear near the unit circle, causing extremely sharp, randomly located spikes in the power spectrum.

5. Algorithms

The most straightforward algorithm for solving the $MEM(p, q)$ problem is to follow the procedure outlined in section 3, building the matrices R_p and $G_{p, q}$, and then calling the RSG driver in the EISPACK library[22] to compute all the generalized eigenvalues λ_i and eigenvectors x_i . If the multiplicity r of the largest eigenvalue is greater than one, then the model order is too high and pole-zero cancellation is a possibility. (In practice we would have to test whether the largest eigenvalues are equal to within a certain tolerance). Either decrease the model order to $MEM(p-r+1, q-r+1)$ and try again, or else find a linear combination of the eigen-

vectors corresponding to λ_{\max} which have the last $r-1$ coefficients all equal to 0. In either case, we set $\hat{\gamma}^2 = 1/\lambda_{\max}$, derive the pole polynomial by scaling the eigenvector corresponding to λ_{\max} , and derive the zero polynomial from equation (3.13).

A faster algorithm relies on the almost-Toeplitz displacement rank 3 structure of $K_\gamma(p, q)$. We start by considering the following set of minimization problems:

$$\hat{a}_n = \min_{\mathbf{a}} \mathbf{a}^H K_\gamma(n, n+q-p) \mathbf{a} \quad (5.1)$$

for $n = 0, \dots, p$, where $\mathbf{a} = (a_{n,n} \cdots a_{1,n} 1)^T$ is a vector constrained to have coefficient $a_{0,n} = 1$, and \hat{a}_n is the value at which the function achieves its minimum.

Also define ϵ_n as the value of this quadratic function at the minimum:

$$\epsilon_n = \hat{a}_n^H K_\gamma(n, n+q-p) \hat{a}_n \quad (5.2)$$

If we choose $\gamma^2 \leq 1/\lambda_{\max}$, then by Lemma 1, $K_\gamma(p, q) \geq 0$. Since by (4.2), $K_\gamma(n, n+q-p)$ is a principle minor of $K_\gamma(p, q)$, we must have $K_\gamma(n, n+q-p) \geq 0$ also. Thus (5.1) involves minimizing a quadratic positive semi-definite function. Differentiating with respect to the real and imaginary parts of $a_{k,n}$ and setting the results to zero, we get a set of linear equations for the minimizing vector \hat{a}_n . Substituting into (5.2) we get a formula for ϵ_n . Combining these gives:

$$K_\gamma(n, n+q-p) \hat{a}_n = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ \epsilon_n \end{pmatrix} \quad (5.3)$$

Because $K_\gamma(n, n+q-p) \geq 0$, we have $\epsilon_n \geq 0$. Note that our *MEM*(p, q) problem is therefore identical to solving the minimization problem (5.1) with gain $\hat{\gamma}^2 = 1/\lambda_{\max}$;

with this gain we will have $\epsilon_p = 0$.

Mullis and Roberts[9] recognized that the displacement rank 3 structure of $K_\gamma(p, q)$ allows us to recursively and quickly compute the a_n, ϵ_n solution in terms of the a_{n-1}, ϵ_{n-1} solution. If the correct gain $\hat{\gamma}^2 = 1/\lambda_{\max}$ were known, therefore, we could run this fast Levinson-style algorithm to compute a_n, ϵ_n for $n = 0, \dots, p$. With this correct gain $\hat{\gamma}^2$, we will get $\epsilon_p = 0$, and the vector a_p will be the pole polynomial solution to the *MEM* (p, q) problem. We will show that the same algorithm can also be used to test the positive definiteness of $K_\gamma(p, q)$. With the guidance of Lemma 1, therefore, we can develop a search algorithm, based on this fast Levinson-style algorithm, to find the value $\hat{\gamma}^2$ which makes $K_{\hat{\gamma}}(p, q)$ exactly positive semi-definite. The *MEM* (p, q) pole polynomial coefficients $\hat{a} = a_p$ will be computed at the same time. At least in theory, this algorithm can also be used to detect when the multiplicity r of the largest eigenvalue is greater than 1; in this case, it automatically gives the unique pole polynomial solution to the *MEM* $(p - r + 1, q - r + 1)$ problem.

The Mullis and Roberts algorithm is motivated by the two recursive definitions of $K_\gamma(p, q)$ in terms of $K_\gamma(p - 1, q - 1)$ given in equations (4.2) and (4.3). In order to solve for a_n efficiently, we will need to calculate auxiliary vectors $d_n = (d_{0,n} \cdots d_{n,n})^T$ and $f_n = (f_{0,n} \cdots f_{n,n})^T$ which satisfy:

$$K_\gamma(n, n + q - p) d_n = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} \quad (5.4)$$

$$\mathbf{K}_\gamma(n, n+q-p) \underline{f}_n = \begin{pmatrix} g^*[q-p] \\ \vdots \\ g^*[n+q-p] \end{pmatrix} \quad (5.5)$$

We will also need to recursively calculate the 2×2 matrix \mathbf{M}_n defined by:

$$\mathbf{M}_n = \begin{pmatrix} \underline{d}_n^H \\ \underline{f}_n^H \end{pmatrix} \mathbf{K}_\gamma(n, n+q-p) \begin{pmatrix} \underline{d}_n & \underline{f}_n \end{pmatrix} \quad (5.6)$$

In Appendix E we show that all these quantities can be recursively computed by the following Levinson-style algorithm:

Mullis-Roberts Algorithm:

Initialization:

$$\bar{R}[n] = \begin{cases} R[n] \\ R[n] - \gamma^2 \sum_{k=0}^{q-p-1} g^*[k]g[k+n] \end{cases} \quad \text{for } n=0, \dots, p$$

$$\underline{a}_0 = \begin{pmatrix} 1 \end{pmatrix}$$

$$\epsilon_0 = \bar{R}(0) - \gamma^2 |g[q-p]|^2$$

$$\underline{d}_0 = \begin{pmatrix} 1/\epsilon_0 \end{pmatrix}$$

$$\underline{f}_0 = \begin{pmatrix} g^*[q-p]/\epsilon_0 \end{pmatrix}$$

$$\mathbf{M}_0 = \frac{1}{\epsilon_0} \begin{pmatrix} 1 \\ g^*[q-p] \end{pmatrix} \begin{pmatrix} 1 & g[q-p] \end{pmatrix}$$

For $n = 1, \dots, p$

$$\phi_n = \sum_{k=0}^{n-1} \bar{R}(n-k) a_{k, n-1}$$

$$\psi_n = \sum_{k=0}^{n-1} g[q-p+n-k]a_{k,n-1}$$

$$a_{k,n} = \begin{cases} 1 & \text{for } k = 0 \\ a_{k,n-1} - \phi_n d_{n-k,n-1} + \gamma^2 \psi_n f_{n-k,n-1} & \text{for } k = 1, \dots, n \end{cases}$$

$$\begin{pmatrix} v_n^* \\ \mu_n^* \end{pmatrix} = M_{n-1} \begin{pmatrix} \phi_n \\ -\gamma^2 \psi_n \end{pmatrix} - \begin{pmatrix} 0 \\ \psi_n \end{pmatrix}$$

$$\epsilon_n = \epsilon_{n-1} - \phi_n v_n + \gamma^2 \psi_n \mu_n$$

IF $n = p$ OR $\epsilon_n \leq 0$ RETURN(n, ϵ_n, a_n)

$$d_{k,n} = d_{k,n-1} - \frac{v_n}{\epsilon_n} a_{n-k,n} \quad \text{for } k = 0, \dots, n$$

$$f_{k,n} = f_{k,n-1} - \frac{\mu_n}{\epsilon_n} a_{n-k,n} \quad \text{for } k = 0, \dots, n$$

$$M_n = M_{n-1} + \frac{1}{\epsilon_n} \begin{pmatrix} v_n^* \\ \mu_n^* \end{pmatrix} \begin{pmatrix} v_n & \mu_n \end{pmatrix}$$

The computation on each pass is approximately $6n + 11$ operations (1 operation \approx 1 multiply + 1 add). Total computation is thus about $3p^2 + 14p$ operations.

Note that we terminate this iteration when $\epsilon_n \leq 0$. The purpose of this is as follows. If $\epsilon_n < 0$ for some n , then by equation (5.2), $K_\gamma(n, n+q-p)$ cannot be positive semi-definite. Since by (4.2) this matrix is a principle minor of $K_\gamma(p, q)$, this last matrix cannot be positive semi-definite either. Thus γ^2 is too large, and we might as well terminate the iteration.

More interestingly, if $\gamma^2 = \hat{\gamma}^2 = 1/\lambda_{\max}$, then $K_{\hat{\gamma}}(n, n+q-p) > 0$ for $0 \leq n < p-r+1$ and $K_{\hat{\gamma}}(p-r+1, q-r+1) \geq 0$ with a null space of dimension one. Equation (5.2) then guarantees that $\epsilon_n > 0$ for $n = 0, \dots, p-r$ and $\epsilon_{p-r+1} = 0$.

Also the minimizing solution a_{p-r+1} will be the *MEM* $(p-r+1, q-r+1)$ pole polynomial solution. Thus if $\gamma^2 = \hat{\gamma}^2$, the test $\epsilon_n \leq 0$ will terminate the algorithm at step $n = p-r+1$, and will return the *MEM* $(p-r+1, q-r+1)$ solution a_{p-r+1} .

This algorithm has some interesting features when $p > q$. In this case, for $n = 0, \dots, p-q-1$ we will have:

$$\underline{d}_n = \frac{1}{\epsilon_n} \begin{pmatrix} 1 \\ a_{1,n} \\ \vdots \\ a_{n,n} \end{pmatrix} \quad (5.7)$$

$$\underline{f}_n = 0$$

$$M_n = \begin{pmatrix} 1/\epsilon_n & 0 \\ 0 & 0 \end{pmatrix}$$

$$\psi_n = \mu_n = 0$$

$$v_n^* = \phi_n/\epsilon_n$$

For the first $p-q-1$ steps, therefore, this algorithm has the same form as Levinson recursion, where a_n plays the role of the forward predictor, $\epsilon_n \underline{d}_n$ plays the role of the backward predictor, ϵ_n is the prediction error, and $-\phi_n/\epsilon_n$ is the reflection coefficient. Furthermore, all these values on the first $p-q-1$ steps will be independent of the value of γ^2 .

This algorithm will work correctly for any value of γ^2 , provided only that $\epsilon_n \neq 0$ for $n = 0, \dots, p-1$. In particular, $K_\gamma(p, q)$ does not have to be positive semi-definite. In fact, the values of ϵ_n can be used to test the positive definiteness of $K_\gamma(p, q)$. Appendix F proves the following:

Theorem 6 Assume that we can find vectors \underline{a}_n and scalars ϵ_n which satisfy (5.3) for $n = 0, \dots, p$. Then:

- A) If $0 \leq \gamma^2 < 1/\lambda_{\max}$, then $\epsilon_0 \geq \epsilon_1 \geq \dots \geq \epsilon_p > 0$. The Mullis-Roberts algorithm will correctly generate all the $\underline{a}_n, \epsilon_n$.
- B) If $\gamma^2 = 1/\lambda_{\max}$, then $\epsilon_0 \geq \dots \geq \epsilon_{p-r} > 0$ and $\epsilon_{p-r+1} = \dots = \epsilon_p = 0$, where r is the multiplicity of the largest eigenvalue λ_{\max} . The Mullis-Roberts algorithm will correctly generate $\underline{a}_n, \epsilon_n$ for $n = 0, \dots, p-r+1$. For $n = p-r+2, \dots, p$ we can set $\underline{a}_n = (0 \ \underline{a}_{n-1}^T)^T$ and $\epsilon_n = 0$.
- C) If $\gamma^2 > 1/\lambda_{\max}$, then at least one ϵ_n is strictly negative.

This theorem suggests a simple binary search algorithm for finding the proper gain value $\hat{\gamma}^2 = 1/\lambda_{\max}$, and for calculating the corresponding pole coefficients \underline{a}_p . The idea is to guess a value for γ^2 , and then run the recursive Mullis-Roberts algorithm. If the ϵ_n are all strictly positive, then $K_\gamma(p, q) > 0$ and the value of γ^2 is too low. If one of the ϵ_n is negative or zero, then $K_\gamma(p, q)$ is not positive definite, and the value of γ^2 is either exactly correct or too high. Adjust the value of γ^2 accordingly, and try again. When the value of γ^2 is known to sufficient precision, we can stop the search, and use the \underline{a}_p vector as the *MEM* (p, q) pole polynomial solution. The zero polynomial coefficients may then be found using (3.13). Putting all this together, we get:

Complete *MEM*(p, q) Algorithm:

Initialization

$$\gamma_L^2 = 0$$

$$\gamma_H^2 = \frac{R[0]}{\sum_{k=0}^q |g[k]|^2}$$

tol = (small) preselected error tolerance

Iterate until $\gamma_H^2 - \gamma_L^2 < tol$

$$\gamma^2 = (\gamma_H^2 + \gamma_L^2)/2$$

$\hat{n}, \epsilon_{\hat{n}}, a_{\hat{n}}$ ← return values from Mullis-Roberts algorithm, using gain γ^2

If $\hat{n} = p$ and $\epsilon_p > 0$

set $\gamma_L^2 = \gamma^2$ and iterate

Else

set $\gamma_H^2 = \gamma^2$ and iterate

When $\gamma_H^2 - \gamma_L^2 < tol$, then DONE

Multiplicity $r = p - \hat{n} + 1$

Calculate the zero polynomial coefficients $b_{k,q-r+1}$ from (3.13)

Return $\gamma_H^2, r, a_{k,p-r+1}, b_{k,q-r+1}$

At each step in this binary search algorithm, γ_L^2 and γ_H^2 represent lower and upper bounds on the feasible range of values of the gain, $\gamma_L^2 < \hat{\gamma}^2 \leq \gamma_H^2$. The initial value $\gamma_L^2 = 0$ comes from Lemma 1. The initial value of γ_H^2 is found by noting that for γ^2 larger than this, the (p, p) element of $K_\gamma(p, q)$ is negative, and thus $K_\gamma(p, q)$ cannot possibly be positive semi-definite. Each loop tests the value at the center of the range, decides whether it is greater than, less than or equal to $\hat{\gamma}^2$. Thus we can cut the feasible range of values for $\hat{\gamma}^2$ in half. Faster procedures than binary search could probably be devised, but binary search has the advantage of achieving a given level of precision for $\hat{\gamma}^2$ after a fixed number of passes.

Due to numerical inaccuracies, it is likely that the preceding algorithm will underestimate the multiplicity of the largest eigenvalue, and thus overestimate the

model order. One potential method of improving this model order identification is to run the binary search algorithm to estimate $\hat{\gamma}^2$. Then call the Mullis-Roberts algorithm again, this time using a gain $\hat{\gamma}^2 + tol2$, where $tol2$ is a small positive number. With the gain set slightly too high, the matrix will be slightly non-positive definite. In particular, if r is the true multiplicity of the largest eigenvalue, then ϵ_{p-r+1} will be slightly negative, and the iteration will stop at step $p-r+1$ despite small numerical errors. Thus the true multiplicity r will be correctly identified, and the Mullis-Roberts algorithm can be run one last time, with the correct gain and model order $MEM(p-r+1, q-r+1)$, to identify the correct low-order ARMA model.

If $p > q$, then the first $p-q+1$ steps of the recursive algorithm will give the same results regardless of the value of γ^2 . To save time, therefore, we need only restart the recursive algorithm at step $n = p-q$ for each new value of γ^2 that we test.

The following theorem, proved in Appendix G, is sometimes useful:

Theorem 7 Assume that $R_p > 0$. Suppose that $\gamma^2 < \hat{\gamma}^2$ so that $\epsilon_0, \dots, \epsilon_p > 0$. Then:

- A) The polynomial $A_p(z) = \sum_{k=0}^p a_{k,p} z^{-k}$ is strictly stable with all roots strictly inside the unit circle.
- B) Form the polynomial $B_q(z)$ by substituting the coefficients of $A_p(z)$ into the recursion (3.13). Then the leading coefficients of the impulse response of $\gamma B_q(z)/A_p(z)$ match $g[0], \dots, g[q]$, but the correlations $R[n]$ of the model are off by:

$$R[n] - \hat{R}[n] = \int_{-\pi}^{\pi} \frac{\epsilon_p}{|\hat{A}_p(e^{j\omega})|^2} \frac{d\omega}{2\pi} \quad (5.8)$$

6. Pisarenko's Method

There is a special case of our $MEM(p, q)$ method which turns out to be identical to Pisarenko's well known method for fitting a harmonic line spectrum to a set of correlations[10]. Assume that there are at least as many zeroes as poles, $q \geq p$, and that the known cepstral values are all zero, $c[n] = 0$ for $n = 1, \dots, q$. Also assume that $R_p > 0$ (this is not necessary to prove the following results, but it allows us to apply our theorems without modification). From the recursive equation (3.11), the leading minimum phase coefficient values must be $g[0] = 1$ and $g[n] = 0$ for $n = 1, \dots, q$. This implies that $G_{p,q}^H G_{p,q} = I$ is an identity matrix, and that the generalized eigenvector problem (3.21) will take the form:

$$\mathbf{x}_i = \lambda_i R_p \mathbf{x}_i \quad (6.1)$$

The solutions for λ_i will be the inverse of the eigenvalues of R_p , and the \mathbf{x}_i will be the corresponding eigenvectors of R_p . Thus $\hat{\gamma}^2 = 1/\lambda_{\max}$ will equal the minimum eigenvalue of R_p , and the $MEM(p, q)$ pole polynomial coefficients \hat{a}_p will be the corresponding eigenvector of R_p . From (3.13), the corresponding zero polynomial will satisfy:

$$\hat{b}_{k,q} = \begin{cases} \hat{a}_{k,p} & \text{for } k = 0, \dots, p \\ 0 & \text{else} \end{cases} \quad (6.2)$$

and thus $\hat{B}_q(z) = \hat{A}_p(z)$. Thus every pole must be canceled by a matching zero, and any extra zeroes will be placed at the origin. Suppose that the multiplicity of the minimum eigenvalue of R_p is r . Combining theorems 2 and 3, at least $p - r + 1$ pole-zero pairs must be exactly on the unit circle.

Since there is complete pole-zero cancellation, the model power spectrum must be flat, $\hat{P}_{p,q}(z) = \hat{\gamma}^2$, and the model correlations are just $\hat{R}[n] = \hat{\gamma}^2 \delta[n]$, where $\delta[0] = 1$ and $\delta[n] = 0$ for $n \neq 0$. Suppose the roots of $\hat{A}_p(z)$ that are on the unit circle are at $\{\rho_i\}$, and that they have multiplicities $\{s_i\}$. Then the correlation matching error formula in Theorem 5 can be used to show that:

$$R[n] = \hat{\gamma}^2 \delta[n] + \sum_i \sum_{j=0}^{s_i-1} \mu_{i,j} n^{2j} \rho_i^n \quad \text{for } n = 0, \dots, p \quad (6.3)$$

The first term represents the correlations of a white noise sequence with variance $\hat{\gamma}^2$. The second term represents the correlations of a sum of at least $p - r + 1$ complex exponentials. In effect, we have modeled the original correlation sequence $R[n]$ as the sum of white noise plus at least $p - r + 1$ complex exponentials whose frequencies are determined by the roots of $\hat{A}_p(z)$ which are on the unit circle. Note that the gains $\mu_{i,j}$ are not determined by our procedure, but would have to be calculated by other means; for example, we could calculate these by factoring $\hat{A}_p(z)$ to find the roots ρ_i , and then solve the set of linear equations (6.3) for the $\mu_{i,j}$.

Finally, note that, except for the calculation of the zero polynomial, $\hat{B}_q(z) = \hat{A}_p(z)$, this procedure for fitting a sum of complex exponentials plus white noise to a given correlation sequence is identical to that suggested by Pisarenko[10]. The interesting point is that we have shown that Pisarenko's method is a special case of our pole-zero MEM method when $q \geq p$ and when the known cepstral coefficients $c[n]$ are all set to zero.

7. Speculation on Non-Minimum Phase $MEM(p,q)$ Models

One troubling aspect of our $MEM(p,q)$ procedure is that the zero polynomial is not guaranteed to be minimum phase, and thus the model is not guaranteed to match the cepstra. Consider, for example, the situation where we are given $R[0]$ and $c[1]$. Our MEM procedure says that the maximum entropy solution should be a single zero model. To calculate it, we compute the leading coefficients of the minimum phase factor $G(z)$ as follows: $g[0]=1$, $g[1]=c[1]$. The $MEM(0,1)$ model is then:

$$\hat{P}_{0,1}(z) = \hat{\gamma}^2 \hat{B}_1(z) \hat{B}_1^*(1/z^*) \quad (7.1)$$

where:

$$\begin{aligned} \hat{B}_1(z) &= 1 + \hat{b}_{1,1} z^{-1} \\ \hat{b}_{1,1} &= g[1] = c[1] \\ \hat{\gamma}^2 &= \frac{R[0]}{1 + |c[1]|^2} \end{aligned} \quad (7.2)$$

The zero polynomial $\hat{B}_1(z)$ has a zero at $-\hat{b}_{1,1} = -c[1]$. For $|c[1]| < 1$, this model will be minimum phase. For $|c[1]| > 1$, the model will not be minimum phase. In fact, to illustrate Theorem 5, for $|c[1]| = |\hat{b}_{1,1}| > 1$, we can factor $\hat{P}_{0,1}(z)$ into its minimum times maximum phase factors as follows:

$$\hat{P}_{0,1}(z) = \gamma^2 |\hat{b}_{1,1}|^2 \left(1 + \frac{1}{\hat{b}_{1,1}} z\right) \left(1 + \frac{1}{\hat{b}_{1,1}^*} z^{-1}\right) \quad (7.3)$$

The minimum phase factor is $(1 + 1/\hat{b}_{1,1} z^{-1})$, and the first cepstral value of the model is $\hat{c}[1] = 1/\hat{b}_{1,1}^*$, which does not equal $c[1] = \hat{b}_{1,1}$. Furthermore, the entropy of this model is $\log(\hat{\gamma}^2 |\hat{b}_{1,1}|^2)$, which is larger than the value we expected, $\log \hat{\gamma}^2$.

This is quite puzzling, since it is easy to find high-order ARMA models which are minimum phase and which match both $R[0]$ and $c[1]$, even when $|c[1]| > 1$. For $n \leq |c[n]| < n+1$, we will need a model with a total of at least $n+1$ poles and zeroes. In fact, it is possible to devise numerical search techniques for finding high order all-pole or all-zero models which match $R[0]$ and $c[1]$ and which have maximum entropy for the given model order. For example, using a numerical search technique sketched in Appendix H, we have found a minimum phase 500 pole model $\gamma^2/A_{500}(z)A_{500}^*(1/z^*)$ which has maximum entropy subject to the constraints that $R[0] = 5.7488221$ and $c[1] = -1.1163325$. The *MEM*(0,1) model for this problem has model coefficients:

$$\begin{aligned} \hat{b}_{1,1} &= -1.1163325 \\ \hat{\gamma}^2 &= 2.5593564 \end{aligned} \tag{7.4}$$

Our AR(500) model has gain $\gamma^2 = 2.5586002$, which is about .04% below the theoretical upper bound of $\hat{\gamma}^2$ above (see Theorem 1). This implies that the entropy of the

	<i>MEM</i> (0,1)	AR(500)
γ^2	2.5593564	2.5586002
$g[0]$	1	1
$g[1]$	-1.11633	-1.11633
$g[2]$	0	.00794
$g[3]$	0	.00768
$R[0]$	5.748822	5.748822
$R[1]$	-2.857093	-2.877325
$R[2]$	0	.0000698
$R[3]$	0	.0000915

Table 7.1 - *MEM*(0,1) versus AR(500) models

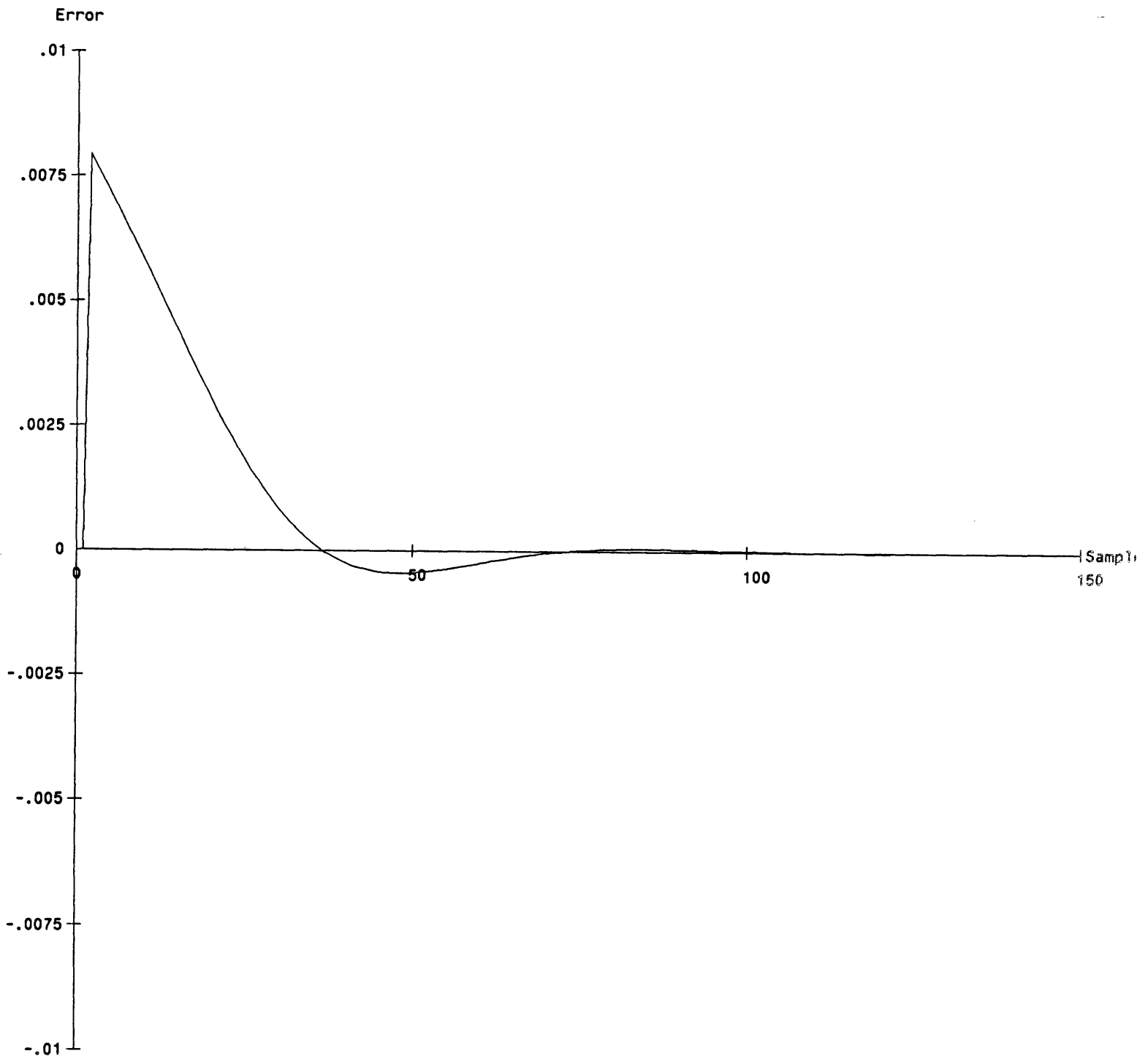


Figure 7.1 - Error Between Impulse Response Coefficients of MEM(0,1) and AR(50) Models

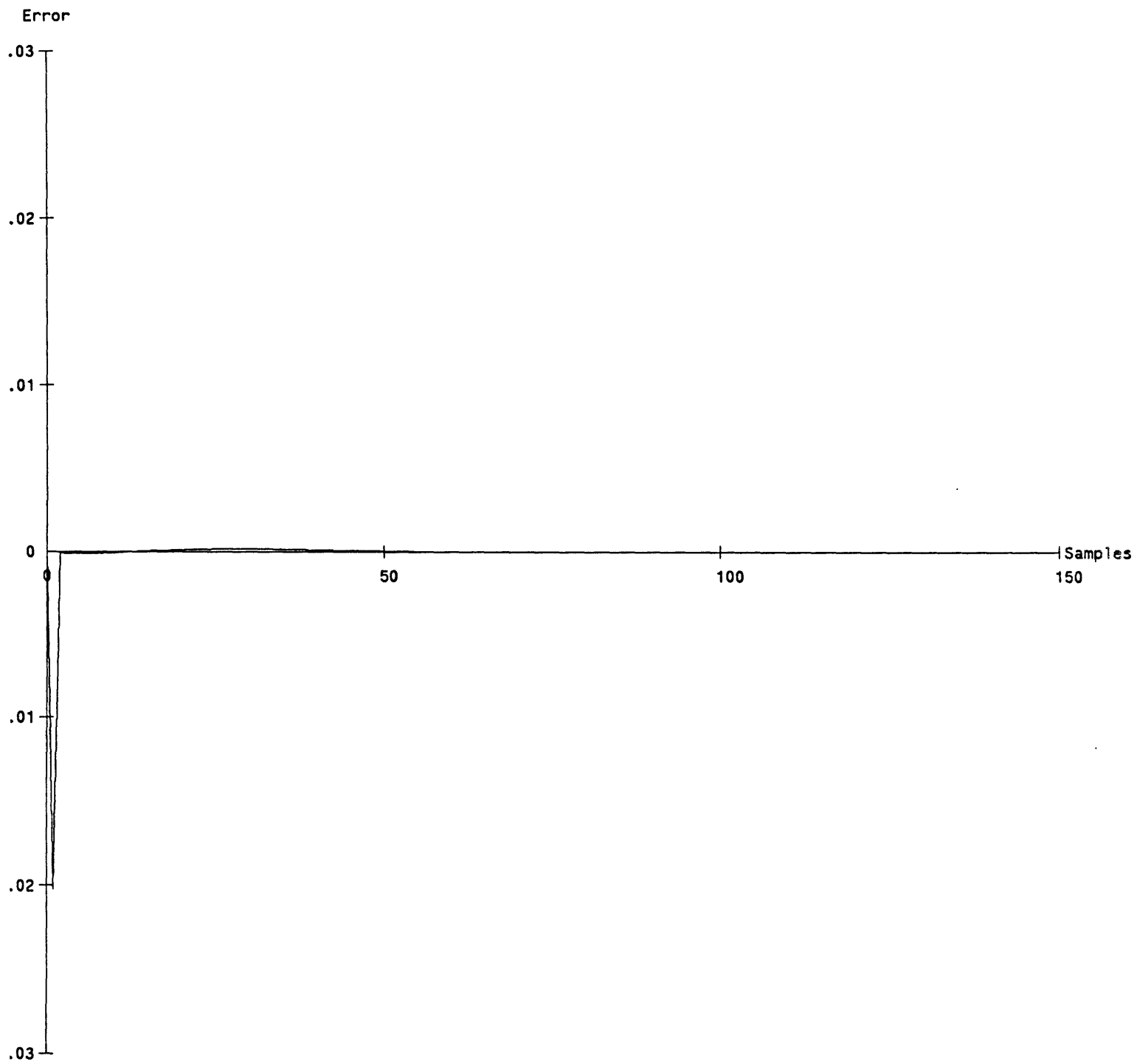


Figure 7.2 - Error Between Correlations of MEM(0,1) and AR(500) Models

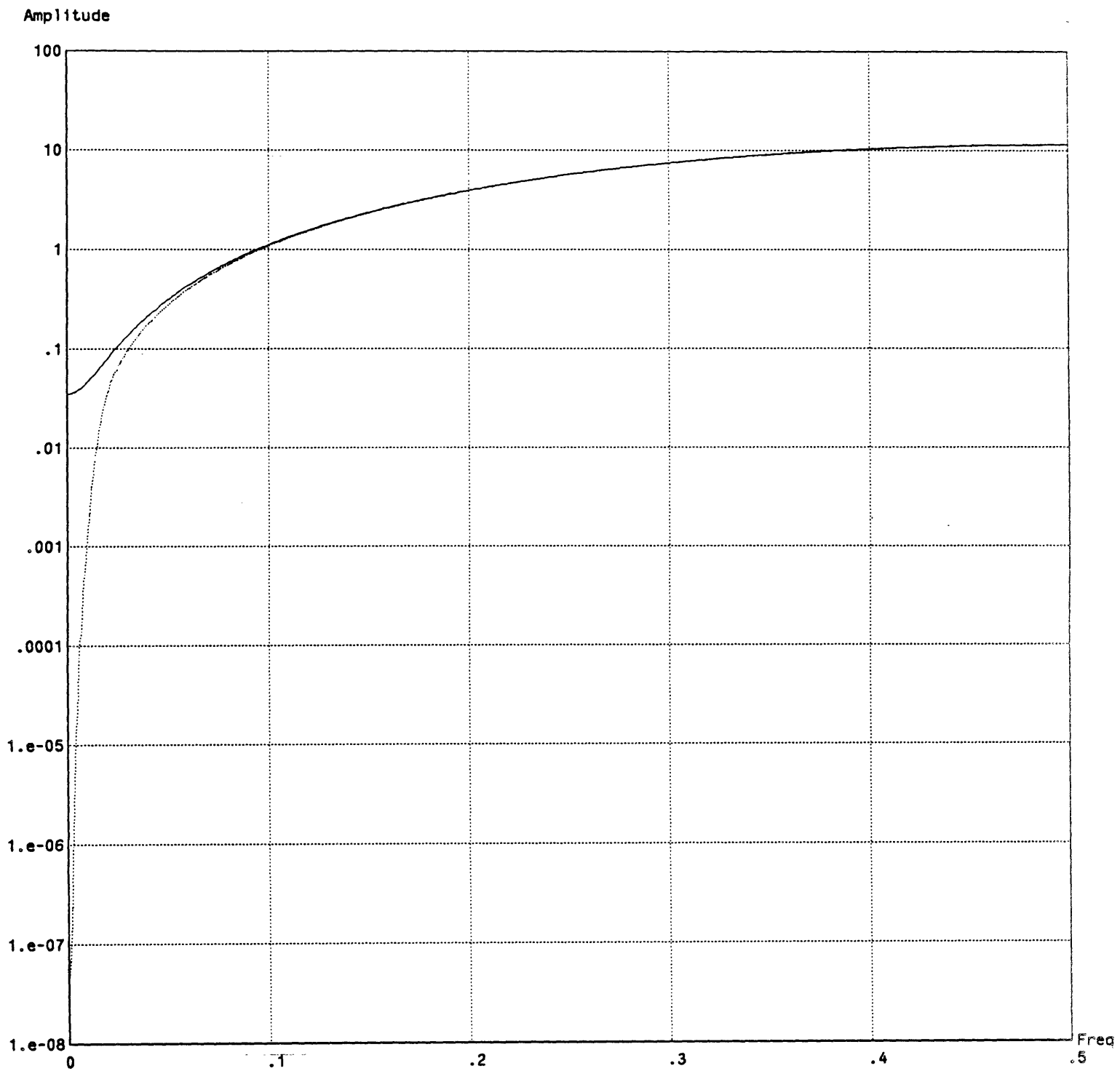


Figure 7.3 - Power Spectra of MEM(0,1) (solid) and AR(500) (dotted) Models

AR (500) model, $H = \log \gamma^2$, is very nearly equal to the maximum achievable entropy. The first 4 leading impulse response coefficients and correlations for the MEM (0,1) and AR (500) models are listed in table 7.1. The remaining impulse response coefficients of the AR (500) model taper to zero. Similarly, the peak AR (500) correlation value for $2 \leq n \leq 501$ is .00016. Figures 7.1 and 7.2 show the errors between the MEM (0,1) and AR (500) impulse responses and correlations. Note that the most significant difference between these models is that the $R[1]$ correlation of the AR (500) model is significantly larger than that of the MEM (0,1) model. In fact, for the AR (500) model, $2R[1]/R[0] > 1$, which is not possible for a single zero model. Figure 7.3 shows the power spectrum of the AR (500) model (dotted line) and the power spectrum of the MEM (0,1) model (solid line). Note that the AR (500) spectrum deviates substantially from the MEM (0,1) spectrum in that there is a deep notch in the AR (500) spectrum at frequency $\omega = 0$. On the basis of this and other similar experiments, we speculate that the following statements may be true:

Speculation 1 Suppose the zero polynomial $\hat{B}_q(z)$ in the MEM ($p-r+1, q-r+1$) problem has zeroes outside the unit circle. Then there exists a sequence of minimum phase polynomials $G_k(z)$ with gains γ_k , such that the impulse response of $\gamma_k G_k(z)$ converges coefficient by coefficient to the impulse response $\hat{\gamma}^2 \hat{B}_{q-r+1} / \hat{A}_{p-r+1}(z)$, and the entropy, $\log \gamma_k^2$, converges from below to the upper bound $\log \hat{\gamma}^2$. Thus the MEM (p, q) solution, though not minimum phase, is a weak limit of a sequence of minimum phase models which asymptotically attain the maximum entropy while meeting all constraints.

Speculation 2 The power spectrum $\gamma_k^2 G_k(z) G_k^*(1/z^*)$ does not converge point by point to the MEM (p, q) power spectrum. Instead, for each zero $z_i = \rho_i e^{j\theta_i}$ which is outside the unit circle, $\rho_i > 1$, there is a notch in the response $\gamma_k^2 \left| G_k(e^{j\omega}) \right|^2$ at frequency $\omega = \theta_i$, and as $k \rightarrow \infty$, the value of the

power spectrum tends to zero at each of these points.

One way to characterize minimum phase polynomials $G_k(z)$ is to compute the unwrapped phase as $z = e^{j\omega}$ transverses the unit circle from $\omega = 0$ to 2π . We define the *winding number* as the difference between the unwrapped phase at $z = e^{j2\pi}$ and at $z = e^{j0}$ divided by 2π . For a minimum phase polynomial, the winding number is zero. For our $MEM(p, q)$ solution, the winding number will equal the negative of the number of zeroes that are outside the unit circle. We speculate that at the notches in the $G_k(z)$, the phase rapidly shifts through a multiple of 2π so that the overall winding number of the polynomial is zero; elsewhere the magnitude and wrapped phase of $G_k(z)$ approximate the $MEM(p, q)$ model.

8. Estimating the Correlations and Cepstra From Time Series Data

To apply our MEM method to estimate the power spectrum of a finite segment of time-series data, it is necessary first to estimate the correlations and cepstra. Suppose we are given complex-valued time series data $x[0], \dots, x[N-1]$, drawn from a stationary Gaussian random process with power spectrum $P_x(\omega)$. The ideal correlations $R[n]$ and cepstra $c[n]$ of the process are defined by formulas (2.2) and (2.4). To estimate these quantities from the given data, we can start by forming an initial estimate of the power spectrum by applying a low pass window $w[n]$ to the data, such as a Hamming window, and then compute the periodogram $I_N(\omega)$:

$$\hat{P}_x(e^{j\omega}) = I_N(\omega) = \frac{1}{N_{eff}} |\bar{X}(e^{j\omega})|^2 \quad (8.1)$$

$$\text{where } \bar{X}(e^{j\omega}) = \sum_{n=0}^{N-1} w[n]x[n]e^{-j\omega n}$$

$$N_{eff} = \sum_{n=0}^{N-1} |w[n]|^2$$

Inverse Fourier transforming the periodogram and the log periodogram then give estimates of the correlation and cepstra respectively:

$$\hat{R}[n] = \int_{-\pi}^{\pi} \hat{P}_x(e^{j\omega}) e^{j\omega n} \frac{d\omega}{2\pi} \quad (8.2)$$

$$\hat{c}[n] = \int_{-\pi}^{\pi} \log \left(\hat{P}_x(e^{j\omega}) \right) e^{j\omega n} \frac{d\omega}{2\pi} \quad (8.3)$$

When the periodogram is used as the power spectrum estimate, and rectangular data windows are used, $w[n] \equiv 1$, then the correlation estimate is equal to:

$$\hat{R}[n] = \frac{1}{N} \sum_{k=0}^{N-1-n} x^*[k]x[k+n] \quad \text{for } n \geq 0 \quad (8.4)$$

In Appendix I we argue that for a rectangular data window, these periodogram-based estimates are asymptotically consistent:

$$E[\hat{R}[n]] \approx R[n] \quad (8.5)$$

$$\text{Var}[\hat{R}[n]] \approx \frac{1}{N} \int_{-\pi}^{\pi} \hat{P}_x^2(e^{j\omega}) \frac{d\omega}{2\pi}$$

$$E[\hat{c}[n]] \approx \begin{cases} c[n] & \text{for } n \neq 0 \\ c[0] - \mu & \text{for } n = 0 \end{cases} \quad (8.6)$$

$$\text{Var}[\hat{c}[n]] \approx \frac{1}{N} \frac{\pi^2}{6}$$

where $\mu \approx .577$ is Euler's constant. Thus the correlation estimates and the cepstral estimates for $n \neq 0$ are asymptotically unbiased as the number of data samples $N \rightarrow \infty$,

and their variance tends toward 0.

One way to improve the estimates of the correlations and cepstra would be to first compute a smooth, consistent power spectrum estimator $\hat{P}_x(e^{j\omega})$, and then use this estimate in formulas (8.2) and (8.3) for $\hat{R}[n]$ and $\hat{c}[n]$. Two basic approaches could be used. A "classical" Blackman-Tukey approach would first estimate L correlations from the data via equation (8.4). It would then multiply by a symmetric window $h[-L], \dots, h[L]$ which has a positive Fourier transform, $H(e^{j\omega}) > 0$. The Fourier Transform of the windowed correlations would then be a positive power spectrum estimate $\hat{P}_x(e^{j\omega})$. This could be plugged into the correlation and cepstral estimation formulas (8.2) and (8.3), and then the $MEM(p, q)$ algorithm could be run. (Note that the correlation estimates will equal our original correlations multiplied by the window $h[n]$). By choosing the correlation window length L much smaller than the data length, $L \ll N$, but much larger than the MEM model order, $L \gg p, q$, we can significantly reduce the variation in the power spectrum estimate, and thus decrease the variation in the cepstral estimates. The tradeoff is that a short window decreases the resolution of the power spectrum estimate, and increases its bias. The resulting correlation and cepstral estimates will thus also be biased. In particular, valleys in the power spectrum tend to be "filled in" by the windowing, and sharp peaks tend to be blurred.

An alternative "modern" approach would start by fitting a high order all-pole MEM model to the first L correlations of the data. We choose L to be much smaller than the number of data points, $L \ll N$, but much larger than the desired $ARMA$

model order, $L \gg p, q$. A Yule-Walker method based on Levinson Recursion could be used; this is identical to our technique for fitting an all-pole $MEM(L,0)$ model to the correlations $\hat{R}[0], \dots, \hat{R}[L]$. This idea has been suggested in numerous contexts for similar problems[23,24]. The resulting all-pole model $\frac{\gamma}{A_L(z)}$ is guaranteed to be minimum phase, and the first L model correlations are guaranteed to match the given correlations. The model $\frac{\gamma}{A_L(z)}$ is thus a good high order approximation to the minimum phase component of the power spectrum. We can use the impulse response of $1/A_L(z)$ as an estimate of $\hat{g}[n]$, then use $\hat{g}[n]$ together with the original correlations $\hat{R}[n]$ to compute the $MEM(p,q)$ model. (It is not necessary to estimate the cepstral values directly, though they could be derived recursively from $\hat{g}[n]$.) By decreasing the model order L of the initial all-pole approximation, we can achieve smoother initial power spectrum estimates and thereby get cepstral estimates with lower variance. However, decreasing the model order also increases the bias of the power spectrum estimate, and the bias of the cepstral estimate.

9. Experimental Results

Figure 9.1 shows the result of using our MEM method with various order models to approximate the impulse response of a minimum phase 8 pole, 4 zero model. The correlations and cepstra were estimated directly from the periodogram via equations (8.2) and (8.3), with a rectangular data window $w[n] \equiv 1$. The dotted lines show the periodogram of the data, while the solid lines show the $MEM(4,0)$, $MEM(6,2)$,

$MEM(8,2)$ and $MEM(8,4)$ approximations. As expected, increasing the model order produces a model spectrum which follows the actual spectrum more and more closely. The $MEM(8,4)$ spectrum exactly matches the data. For model orders higher than this, the algorithm either detects that the largest eigenvalue has multiplicity greater than 1 and automatically decreases the model order to its correct value, or else it places extra canceling pole-zero pairs exactly on the unit circle, as predicted by Theorem 3. In either case, the model spectra exactly match the ideal spectrum.

Unfortunately, our MEM method does not perform as well when we must estimate the correlations and cepstra from a finite segment of time-series data. Figure 9.2, for example, shows the result of applying the method to data derived by passing an impulse train with period 91 through the same 8 pole, 4 zero filter as used in figure 9.1. 12 different datasets were used, each 800 points long, each differing solely in the initial phase of the impulse train excitation. Three different methods were used to generate an initial power spectrum estimate from the 12 Hamming-windowed datasets. In figure 9.2a, we used the periodogram. In figure 9.2b we applied a triangular lag window to the first $L=50$ correlations. In figure 9.2c we used an initial $MEM(50,0)$ approximation. Correlations and cepstra were computed via formulas (8.2) and (8.3) from these initial power spectra. $MEM(10,6)$ models were fit to the estimated correlations and cepstra from each dataset, and the resulting spectra were drawn in solid lines, superimposed on the periodogram. We deliberately used a model with an extra pole and zero pair in order to demonstrate the behavior when the model order is set too high.

Note that the periodogram-based method in figure 9.2a shows 5 strong peaks instead of 4, with resonances that are far too sharp. Note also the large variance in the spectral models for the 12 datasets. 4 of the models are not minimum phase. In figure 9.2b, using a 50 point triangular window on the correlations before estimating the correlations and cepstra gives MEM models which follow the 4 peaks more robustly. The valleys of the spectra, however, show the characteristic "filling in" due to the initial windowing operation. All of the models are minimum phase. If we substitute the original correlations for the windowed correlations, but use the cepstra from the log windowed periodogram, then we would get MEM spectra with 5 strong, excessively resonant peaks. (This is not shown). The spectra have a large amount of variance, and 11 of them are not minimum phase. This suggests that it is important to start with correlation and cepstral estimates derived from the same initial power spectrum estimate.

Using an initial MEM(50,0) model to estimate the correlations and cepstra in figure 9.2c gives the best results. The models have low variance, and all conform closely to the original filter shape. Note, however, the peculiar glitch in many of the models near a frequency of 0.27. This is caused by a nearly canceling pole-zero pair located virtually on the unit circle. None of the models are minimum phase. The culprit appears to be related to Theorem 3. In all 12 datasets, the multiplicity of the largest eigenvalue is estimated as $r = 1$. Given more poles and zeros than necessary, however, the method tries to locate the extra poles and zeroes someplace where they will nearly cancel. Unfortunately, by theorem 3, if the multiplicity $r = 1$, then the only place

canceling pole-zero pairs can be located is exactly on the unit circle. Since the extra poles and zeroes do not exactly cancel, they are only placed near each other and near the unit circle. The result is a sharp, highly variable "glitch" in the power spectrum.

Figure 9.2d shows the results when the correct model order, $MEM(8,4)$ is used, with an initial $MEM(40,0)$ model to estimate the correlations and cepstra. Note that the 12 model spectra are quite accurate, except for some variance in the depth of the high frequency null in the spectrum. All are minimum phase.

Figure 9.3 shows the result of applying the method to a similar problem, where white Gaussian noise is filtered through the same 8 pole, 4 zero model to produce a colored Gaussian signal. 12 different datasets were generated, each 800 points long. Exactly the same processing was used, except that the model order was increased to $MEM(12,8)$. Figure 9.3a was generated using an initial power spectrum estimate equal to the periodogram. Figure 9.3b used a 60 point triangular lag window, while figure 9.3c used an initial $MEM(60,0)$ model. The periodogram-based and $MEM(60,0)$ based models are very similar; all show some variance in the spectral envelope, together with 2 sharp glitches caused by the 4 extra poles and zeroes being located in nearly canceling pairs near the unit circle. The triangular-window based models have much less variance, but show the characteristic "filling in" of the valleys. Figure 9.3d shows the models generated with the correct $MEM(8,4)$ model order, using the correlations and cepstra derived from the periodogram. Note that the correct spectral envelope is captured in all 12 datasets.

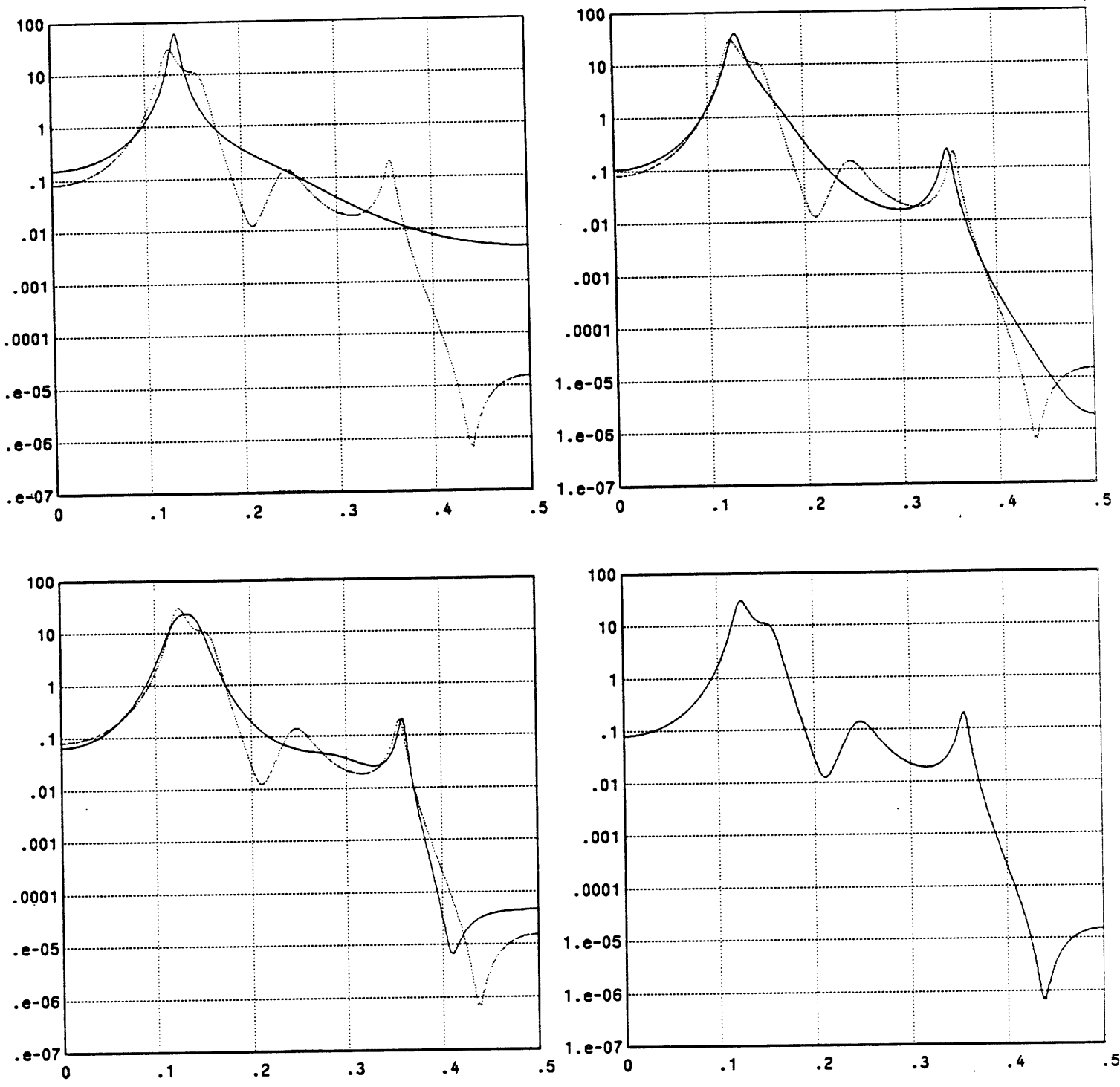
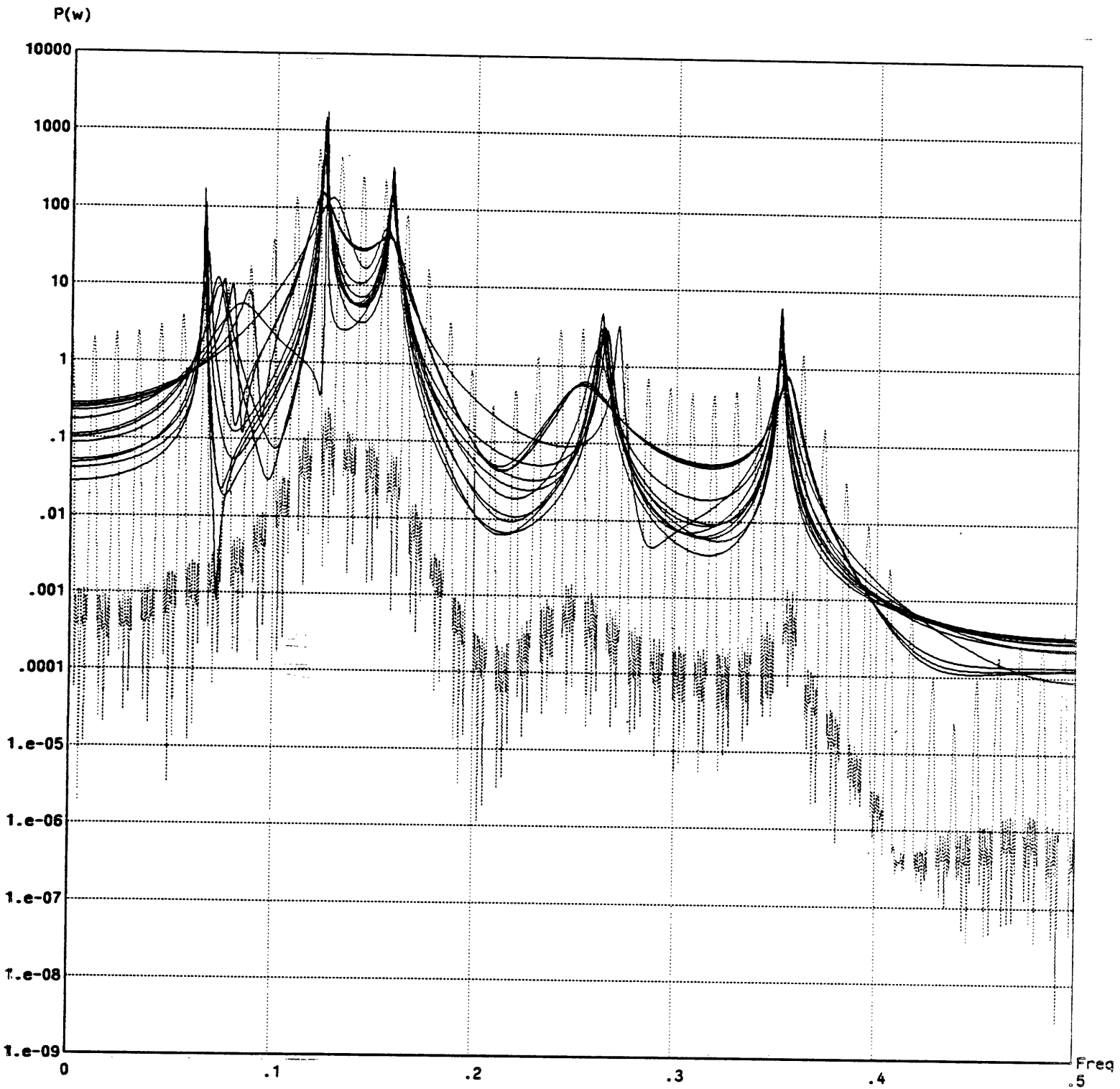
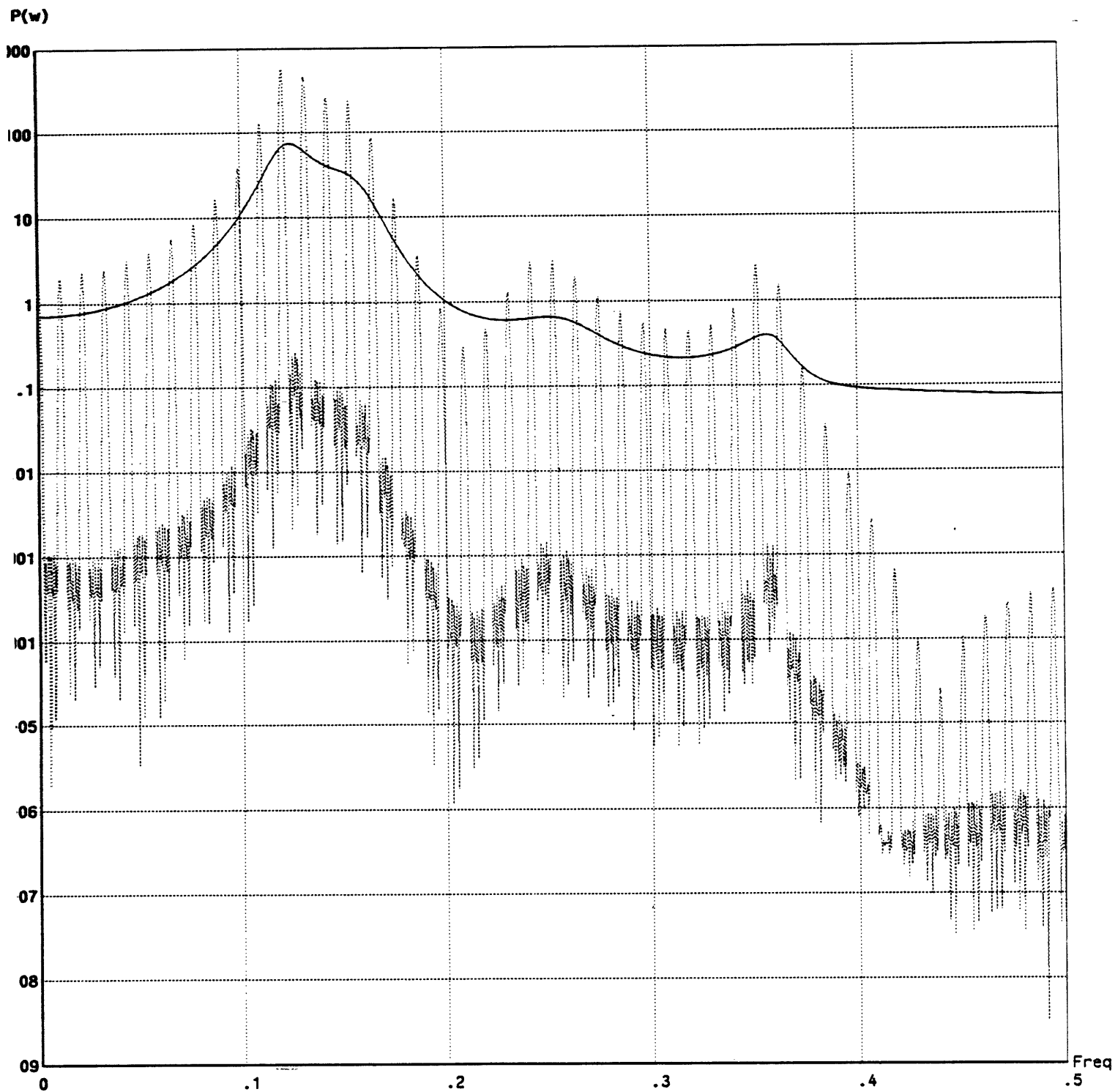


Figure 9.1 - MEM(4,0), MEM(6,2), MEM(8,2), MEM(8,4) (solid) and original spectrum (dotted)



**Figure 9.2a - Impulse Train Excited Filter, Periodogram Based MEM(10,6)
Model (solid), Periodogram (dotted)**



**Figure 9.2b - Impulse Train Excited Filter, Triangular Window Based
MEM(10,6) Model (solid), Periodogram (dotted)**

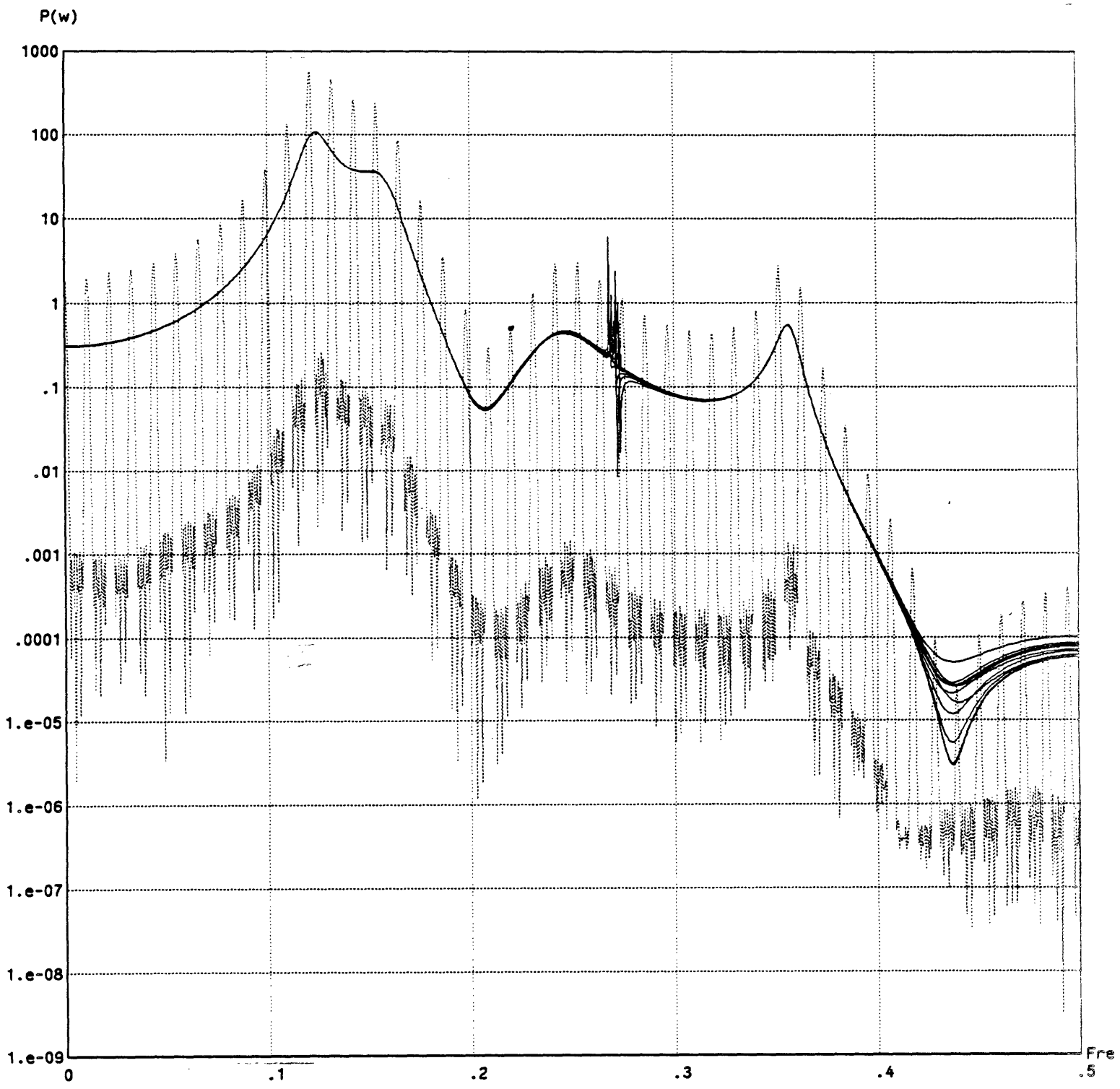
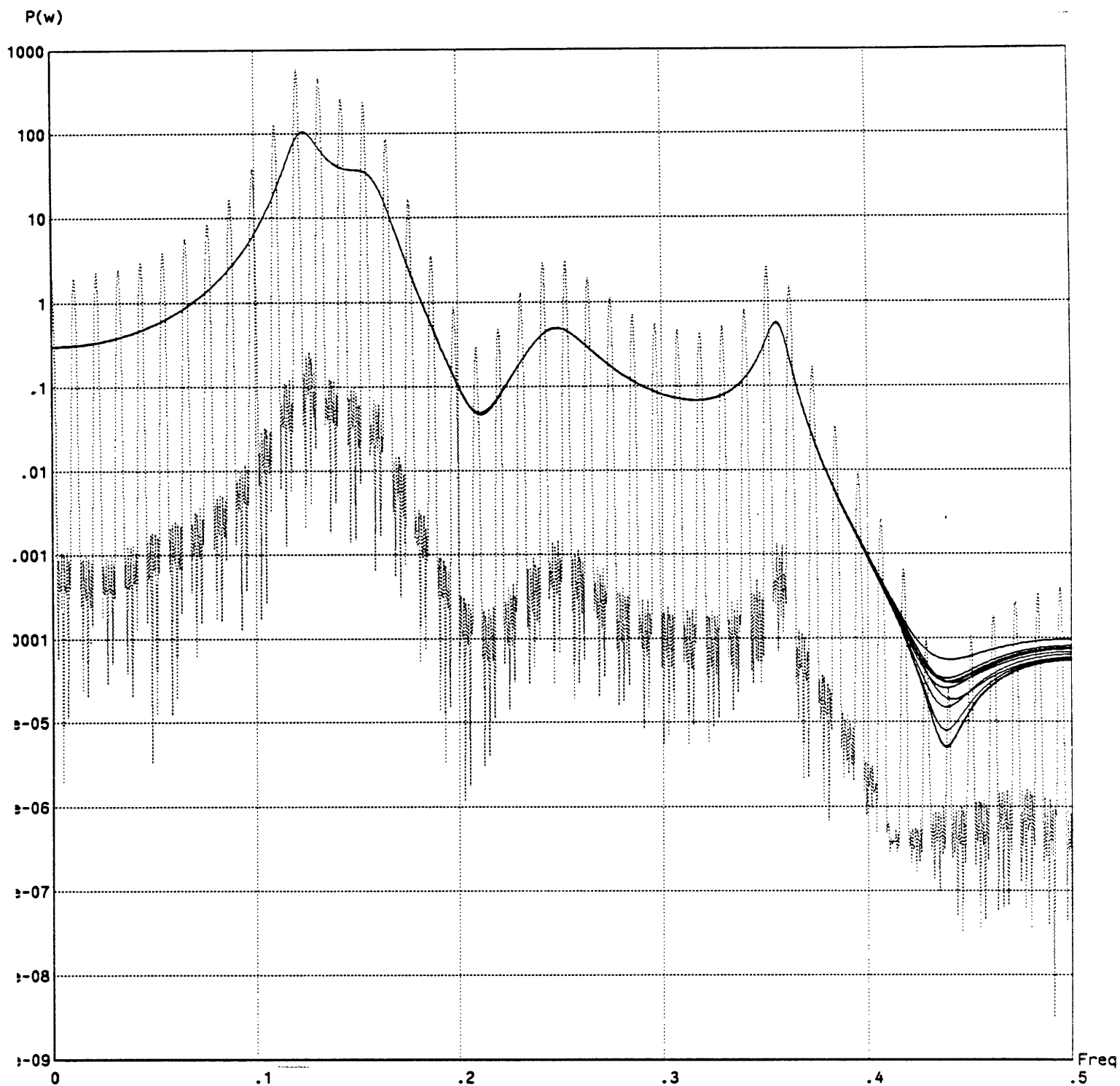
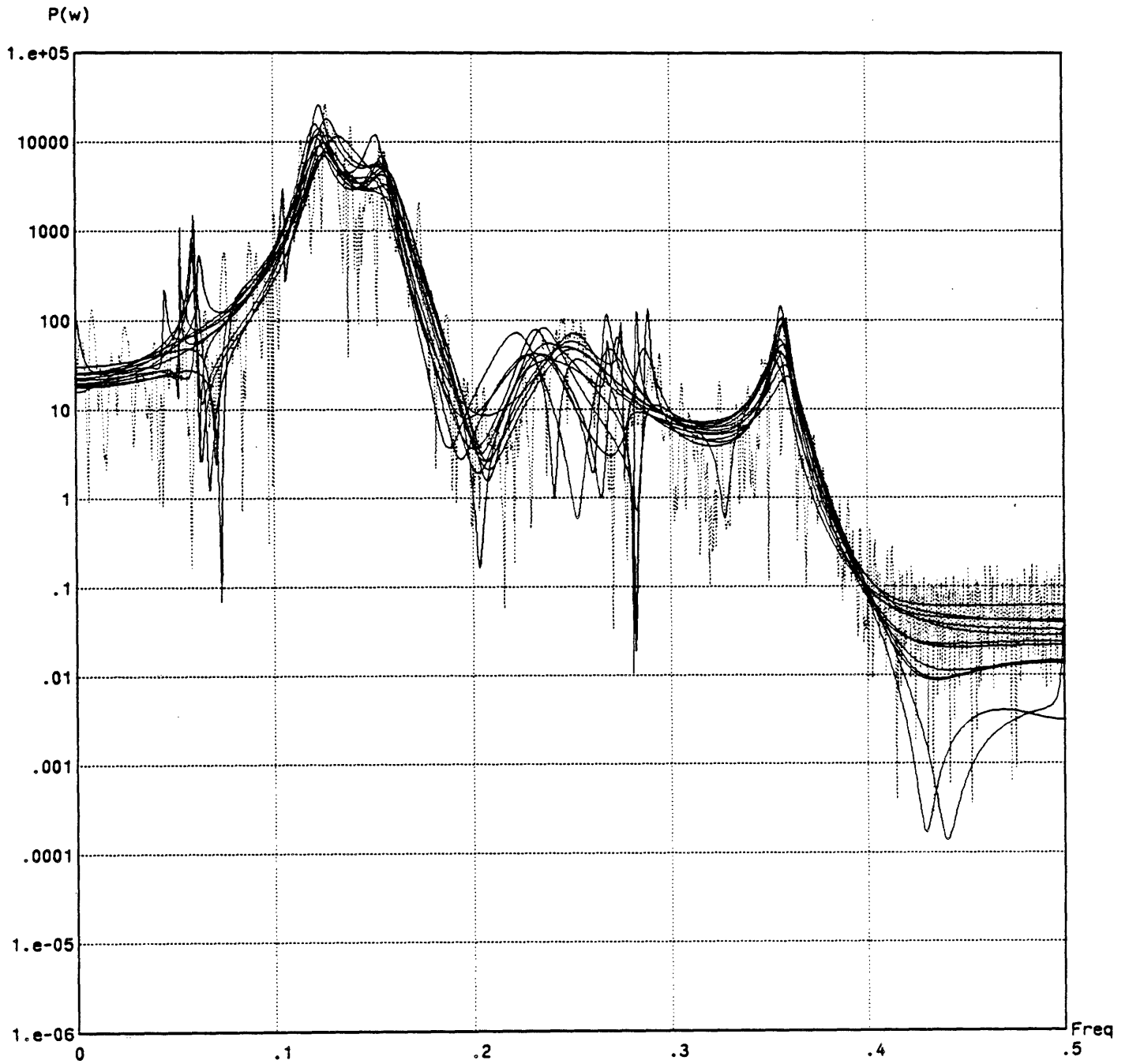


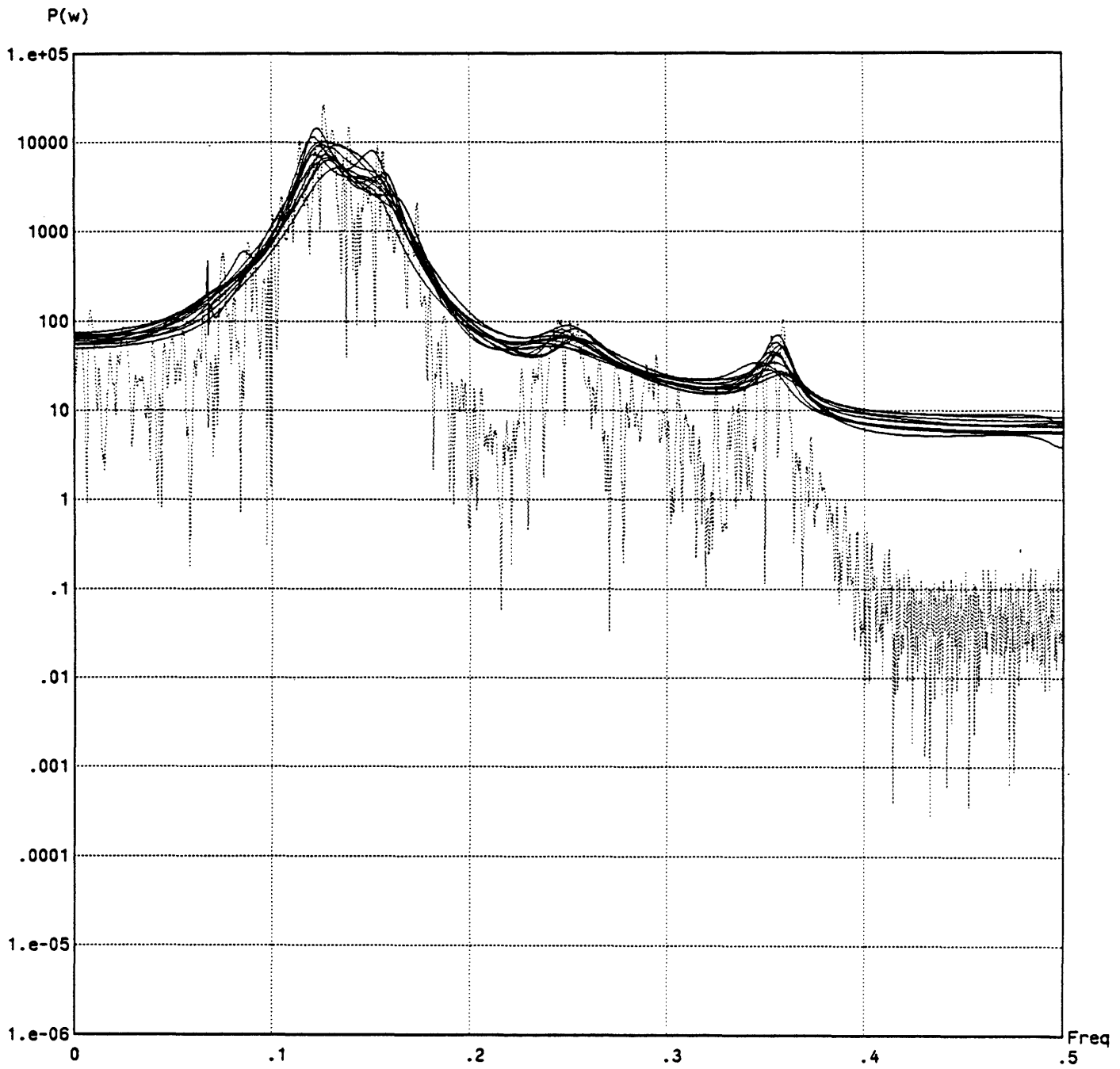
Figure 9.2c - Impulse Train Excited Filter, AR(50) Based MEM(10,6) Model (solid), Periodogram (dotted)



**Figure 9.2d - Impulse Train Excited Filter, AR(40) Based MEM(8,4) Model
(solid), Periodogram (dotted)**



**Figure 9.3a - Colored Gaussian Noise, Periodogram Based MEM(12,8)
Model (solid), Periodogram (dotted)**



**Figure 9.3b - Colored Gaussian Noise, Triangular Window Based
MEM(12,8) Model (solid), Periodogram (dotted)**

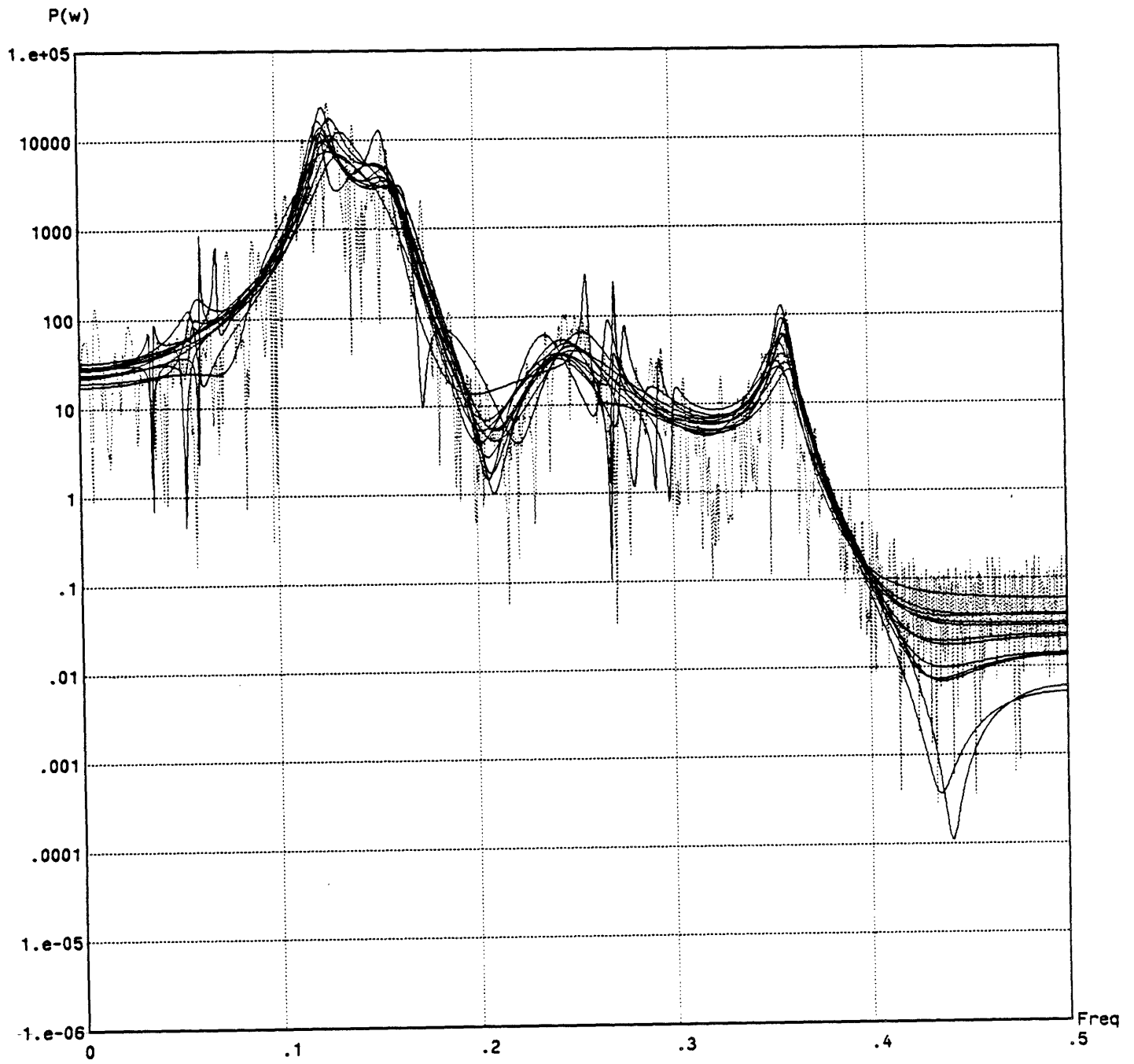


Figure 9.3c - Colored Gaussian Noise, AR(50) Based MEM(12,8) Model (solid), Periodogram (dotted)

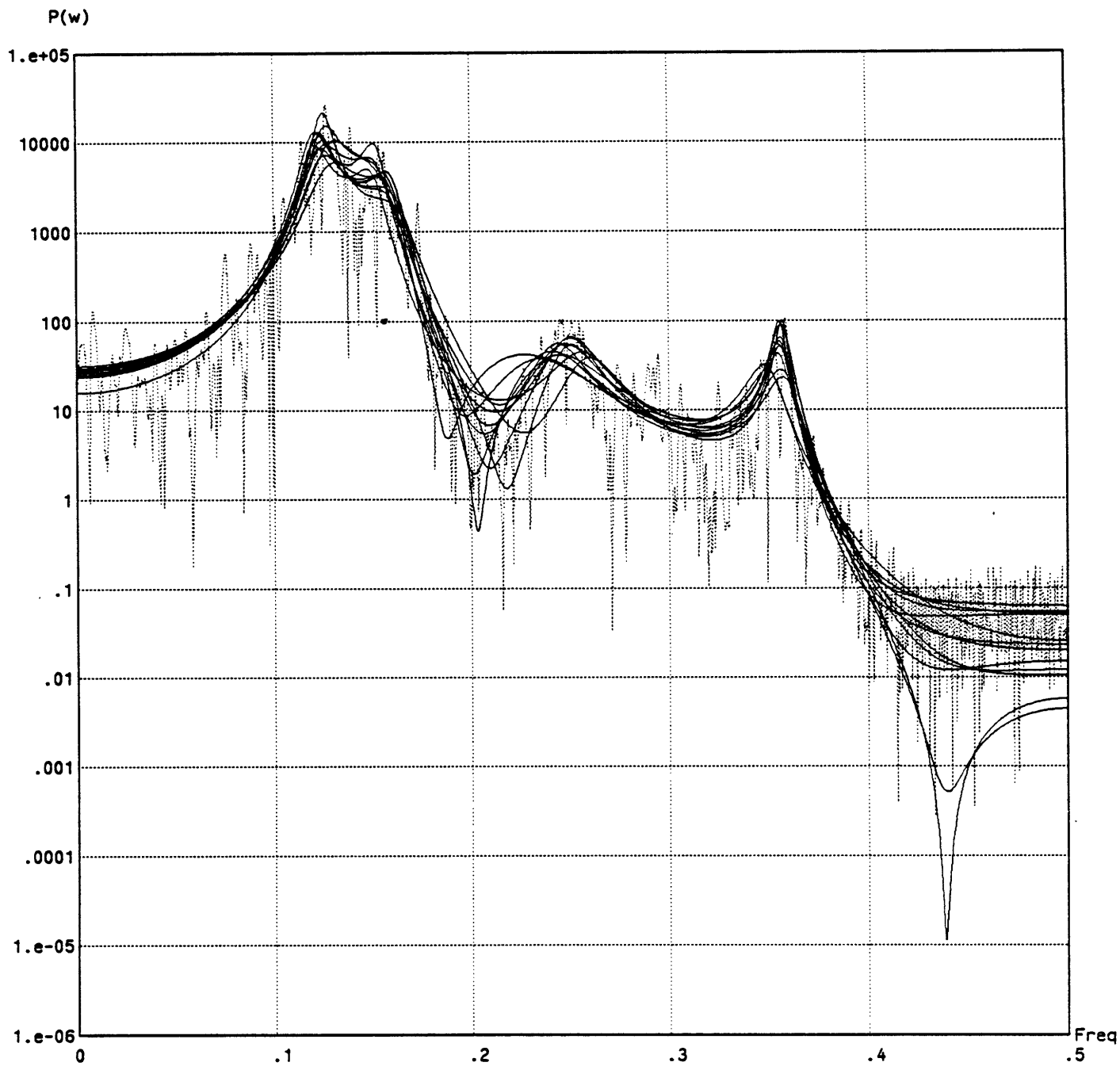


Figure 9.3d - Colored Gaussian Noise, AR(40) Based MEM(8,4) Model (solid), Periodogram (dotted)

In general, resolution in the model power spectrum appears to be controlled by the number of poles used in the model; adding more zeroes tends to improve the shape of the spectral peaks. In practice it is rare for the method to detect multiple eigenvalues, even when the model order is set relatively high. Non-minimum phase models do not appear to have significantly different shape than minimum phase models generated from datasets with the same stochastic behavior. When modeling data formed from sinusoids in white Gaussian noise, the method tends to locate nearly canceling pole-zero pairs near the correct frequencies; unfortunately, these result in peaks whose position and shape appear to be highly variable. The worst problem, in general, is the presence of sharp glitches in the power spectrum when the model order is high. Further work is needed to devise an accurate model order estimation procedure, or to eliminate this behavior.

10. Conclusions and Further Work

We have shown that the solution to the Maximum Entropy spectral estimation problem, subject to constraints on the first $p + 1$ correlations and first q cepstral values, is an ARMA model with p poles and q zeroes. To calculate the parameters of this model, we converted the Maximum Entropy problem into an equivalent generalized eigenvalue/eigenvector problem. The inverse of the maximum eigenvalue in this problem is the model gain; it is also the exponential of the entropy of the model. The corresponding eigenvector contains the coefficients of the pole polynomial. The zero polynomial coefficients are computed by a simple recursion. The solution is unique,

except for possible pole-zero cancellation. The model is also stable, with all uncanceled poles inside the unit circle. If all the zeroes are also inside the unit circle, then the computed model matches all the given correlation and cepstral values and it is the Maximum Entropy solution. If the zero polynomial is not minimum phase, however, then it does not match the cepstra, and there does not exist any minimum phase model which matches the given constraints, has a finite and non-zero power spectrum, and also achieves the maximum possible entropy. We speculate, however, that there exist minimum phase models which match the correlations and cepstra, whose impulse responses match that of the $MEM(p, q)$ model arbitrarily closely, and whose entropy is arbitrarily close to the theoretical upper bound given by the $MEM(p, q)$ method.

Several issues regarding this algorithm require considerable work. The issue of non-minimum phase models needs to be resolved more cleanly; in particular, a better explanation would be helpful for why the non-minimum phase models seem to yield spectra which are about as accurate as the minimum phase models. Better methods are needed to estimate the correlations and cepstra more carefully from the given data, or to modify the algorithm to give more accurate estimates from short segments of data. Model order estimation procedures are needed. These would probably be based on monitoring the value of the gain $\hat{\gamma}^2$ or the error ϵ_n as the model order is increased, and picking a model order near the knee in this curve. Modifications to improve the estimation of sinusoids in noise would be helpful. Most importantly, the problem of nearly canceling pole-zero pairs appearing near the unit circle must be resolved. Despite these difficulties, however, this new pole-zero MEM method is quite elegant, and seems quite promising.

Appendix A - Proofs of Lemma 1 and Theorem 1

Proof of Lemma 1

Assume that $R_p > 0$. Then we can always factor R_p in the form $R_p = Q^H Q$, where Q is an invertible matrix. (For example, Q could be the upper triangular factor computed by a Choleski decomposition [18]). Let Q^{-H} be the inverse of Q^H . Then:

$$Q^{-H} K_\gamma(p, q) Q^{-1} = I - \gamma^2 Q^{-H} G_{p,q}^H G_{p,q} Q^{-1} \quad (\text{A.1})$$

Note that for all \mathbf{x} :

$$\mathbf{x}^H Q^{-H} G_{p,q}^H G_{p,q} Q^{-1} \mathbf{x} = \mathbf{y}^H \mathbf{y} \geq 0 \quad (\text{A.2})$$

where $\mathbf{y} = G_{p,q} Q^{-1} \mathbf{x}$. Thus the matrix $Q^{-H} G_{p,q}^H G_{p,q} Q^{-1}$ is positive semi-definite, and it will have a complete set of orthonormal eigenvectors $\mathbf{x}_0, \dots, \mathbf{x}_p$ and corresponding eigenvalues $\lambda_0, \dots, \lambda_p$, all of which are non-negative:

$$Q^{-H} G_{p,q}^H G_{p,q} Q^{-1} \mathbf{x}_i = \lambda_i \mathbf{x}_i \quad \text{for } i=0, \dots, p \quad (\text{A.3})$$

where $\lambda_i \geq 0$

$$\mathbf{x}_i^H \mathbf{x}_j = \delta_{i,j}$$

where $\delta_{i,i} = 1$ and $\delta_{i,j} = 0$ for $i \neq j$. Let us order the eigenvalues so that λ_0 is the largest and λ_p is the smallest, and let us define $\lambda_{\max} = \lambda_0$.

We next show the connection between these eigenvalues and eigenvectors, and the generalized eigenvalue problem (3.21). Suppose λ_i, \mathbf{x}_i is an eigenvalue, eigenvector pair of the matrix $Q^{-H} G_{p,q}^H G_{p,q} Q^{-1}$. Let $\mathbf{y}_i = Q^{-1} \mathbf{x}_i$. Then:

$$G_{p,q}^H G_{p,q} \mathbf{y}_i = Q^H \left(Q^{-H} G_{p,q}^H G_{p,q} Q^{-1} \right) \mathbf{x}_i \quad (\text{A.4})$$

$$\begin{aligned}
&= \lambda_i Q^H \mathbf{x}_i \\
&= \lambda_i Q^H Q \mathbf{v}_i \\
&= \lambda_i \mathbf{R}_p \mathbf{v}_i
\end{aligned}$$

Thus λ_i, \mathbf{v}_i is a generalized eigenvalue, eigenvector pair solving (3.21). Conversely, it is easy to show that if λ_i, \mathbf{v}_i solves (3.21), then λ_i and $\mathbf{x}_i = Q \mathbf{v}_i$ is an eigenvalue, eigenvector pair of the matrix $Q^{-H} G_{p,q}^H G_{p,q} Q^{-1}$. Thus the generalized eigenvalue, eigenvector problem (3.21) has the same $(p+1)$ non-negative eigenvalues as does the matrix $Q^{-H} G_{p,q}^H G_{p,q} Q^{-1}$. Also, the corresponding eigenvectors $\mathbf{v}_0, \dots, \mathbf{v}_p$ are linearly related to the eigenvectors $\mathbf{x}_0, \dots, \mathbf{x}_p$ by $\mathbf{v}_i = Q^{-1} \mathbf{x}_i$. Therefore they are linearly independent and satisfy:

$$\mathbf{v}_i^H \mathbf{R}_p \mathbf{v}_j = \mathbf{x}_i^H \mathbf{x}_j = \delta_{i,j} \quad (\text{A.5})$$

Now suppose that the maximum eigenvalue λ_{\max} has multiplicity r , so that $\lambda_{\max} = \lambda_0 = \dots = \lambda_{r-1}$. For model orders $q \geq 0$, the matrix $G_{p,q}$ has at least one non-zero element. Since Q is invertible, the number of non-zero eigenvalues of $Q^{-H} G_{p,q}^H G_{p,q} Q^{-1}$ will equal the rank of $G_{p,q}$; since this is at least one, $\lambda_{\max} > 0$.

Let \mathbf{v} be an arbitrary non-zero vector. Because the $\mathbf{v}_0, \dots, \mathbf{v}_p$ form a basis, we can always write \mathbf{v} as a linear combination of these eigenvectors:

$$\mathbf{v} = \sum_{i=0}^p \alpha_i \mathbf{v}_i \quad (\text{A.6})$$

where at least one α_i is non-zero. But then, using $\mathbf{v}_i = Q^{-1} \mathbf{x}_i$ together with (A.1) and (A.3):

$$\begin{aligned}
\mathbf{y}^H \mathbf{K}_\gamma(p, q) \mathbf{y} &= \sum_{i=0}^p \sum_{j=0}^p \alpha_i \alpha_j^* \mathbf{x}_i^H \mathbf{Q}^{-H} \mathbf{K}_\gamma(p, q) \mathbf{Q}^{-1} \mathbf{x}_j \\
&= \sum_{i=0}^p \sum_{j=0}^p \alpha_i \alpha_j^* \mathbf{x}_j^H (1 - \gamma^2 \lambda_j) \mathbf{x}_j \\
&= \sum_{i=0}^p |\alpha_i|^2 (1 - \gamma^2 \lambda_i)
\end{aligned} \tag{A.7}$$

For $0 \leq \gamma^2 < 1/\lambda_{\max}$, every term on the right hand side of (A.7) will be non-negative, and at least one will be strictly positive. Thus $\mathbf{y}^H \mathbf{K}_\gamma(p, q) \mathbf{y} > 0$ for all $\mathbf{y} \neq \mathbf{0}$, and $\mathbf{K}_\gamma(p, q)$ will be strictly positive definite. For $\gamma^2 = 1/\lambda_{\max}$, terms 0 through $r-1$ will be zero, so that:

$$\mathbf{y}^H \mathbf{K}_\gamma(p, q) \mathbf{y} = \sum_{i=r}^p |\alpha_i|^2 (1 - \lambda_i / \lambda_{\max}) \tag{A.8}$$

Thus $\mathbf{K}_\gamma(p, q)$ will be positive semi-definite, and its null space will be spanned by the vectors $\mathbf{y}_0, \dots, \mathbf{y}_{r-1}$. Since these vectors are linearly independent, the null space of $\mathbf{K}_\gamma(p, q)$ will have dimension r . For $\gamma^2 > 1/\lambda_{\max}$, we will have $\mathbf{y}_0^H \mathbf{K}_\gamma(p, q) \mathbf{y}_0 < 0$, and so $\mathbf{K}_\gamma(p, q)$ will not be positive semi-definite.

Proof of Theorem 1

If there exists some minimum phase model γ , $G(z)$ which matches the given correlations and impulse response coefficients, then section 3 shows that $\mathbf{K}_\gamma(p, q)$ must be positive semi-definite. By Lemma 1, this requires $0 \leq \gamma^2 \leq 1/\lambda_{\max}$. By section 3, the entropy of the model is $H = \log \gamma^2$, and thus the entropy must be bounded above by $H < \log \hat{\gamma}^2 = \log 1/\lambda_{\max}$.

For the $MEM(p, q)$ problem, we must choose γ^2 so that $K_\gamma(p, q)$ is positive semi-definite with a non-trivial null space. By Lemma 1, the only possible choice is

$$\hat{\gamma}^2 = 1/\lambda_{\max}.$$

Appendix B - Proof of Theorem 2

Assume that $R_p > 0$ and that λ_{\max} has multiplicity r . For notational convenience, we will call the n^{th} order polynomial $V_n(z) = \sum_{k=0}^n v_k z^{-k}$ a member of the null space of $K_\gamma(n, m)$ if the vector of coefficients $\underline{v} = (v_n \cdots v_0)^T$ is a member of the null space. (For convenience, we drop the model order index from the coefficients.) Also, let us define $K_\gamma(n, m) \equiv R_n$ if $m < 0$ but $n \geq 0$.

In Appendix A we proved that for $\hat{\gamma}^2 = 1/\lambda_{\max}$, then $K_{\hat{\gamma}}(p, q) \geq 0$, and its null space has dimension r . This null space is spanned by the generalized eigenvectors $\underline{v}_0, \dots, \underline{v}_{r-1}$. Let us look for vectors \underline{v} in the null space of $K_{\hat{\gamma}}(p, q)$ whose last coefficient is zero, $v_0 = 0$. This represents only one additional linear constraint on the values of v_0, \dots, v_r ; thus there will be at least an $r-1$ dimensional linear subspace of vectors $\underline{v}' = (v_p \cdots v_1 0)^T$ which are elements of the null space of $K_{\hat{\gamma}}(p, q)$. But equation (4.2) implies that if $K_{\hat{\gamma}}(p, q)\underline{v}' = \underline{0}$, then:

$$K_{\hat{\gamma}}(p-1, q-1) \begin{pmatrix} v_p \\ \vdots \\ v_1 \end{pmatrix} = \underline{0} \quad (\text{B.1})$$

Thus $K_{\hat{\gamma}}(p-1, q-1)$ must have a null space with dimension at least equal to $r-1$. Applying this argument recursively for $s = 1, \dots, r-1$, we can show that $K_{\hat{\gamma}}(p-s, q-s)$ must have a null space with dimension of at least $r-s$.

Now let \hat{s} be the largest integer such that $K_{\hat{\gamma}}(p-\hat{s}, q-\hat{s})$ has a non-trivial null space. By the reasoning above, $\hat{s} \geq r-1$. Let $\hat{A}_{p-\hat{s}}(z)$ be any non-zero member of

the null space of $K_{\hat{\gamma}}(p-\hat{s}, q-\hat{s})$:

$$K_{\hat{\gamma}}(p-\hat{s}, q-\hat{s}) \begin{pmatrix} \hat{a}_{p-\hat{s}} \\ \vdots \\ \hat{a}_0 \end{pmatrix} = \mathbf{0} \quad (\text{B.2})$$

First we show that $\hat{a}_0 \neq 0$. Suppose this were not true. If $\hat{s} = p$, then we would have $\hat{A}_{p-\hat{s}}(z) = 0$, which would contradict our definition that $\hat{A}_{p-\hat{s}}(z)$ is a non-trivial member of the null space of $K_{\hat{\gamma}}(p-\hat{s}, q-\hat{s})$. For $\hat{s} < p$, using equation (4.2) we would have:

$$\begin{aligned} 0 &= (\hat{a}_{p-\hat{s}}^* \cdots \hat{a}_1^* 0) K_{\hat{\gamma}}(p-\hat{s}, q-\hat{s}) \begin{pmatrix} \hat{a}_{p-\hat{s}} \\ \vdots \\ \hat{a}_1 \\ 0 \end{pmatrix} \\ &= (\hat{a}_{p-\hat{s}}^* \cdots \hat{a}_1^*) K_{\hat{\gamma}}(p-\hat{s}-1, q-\hat{s}-1) \begin{pmatrix} \hat{a}_{p-\hat{s}} \\ \vdots \\ \hat{a}_1 \end{pmatrix} \end{aligned} \quad (\text{B.3})$$

But since $K_{\hat{\gamma}}(p, q) \geq 0$, and since $K_{\hat{\gamma}}(p-\hat{s}-1, q-\hat{s}-1)$ is just the $p-\hat{s}-1$ order principle minor of $K_{\hat{\gamma}}(p, q)$, we must have $K_{\hat{\gamma}}(p-\hat{s}-1, q-\hat{s}-1) \geq 0$. By definition of \hat{s} , however, $K_{\hat{\gamma}}(p-\hat{s}-1, q-\hat{s}-1)$ can have only a trivial null space and so this matrix must be strictly positive definite. But then (B.3) implies that $\hat{a}_n = 0$ for $n = 1, \dots, p-\hat{s}$. This, however, contradicts our original assumption that $\hat{A}_{p-\hat{s}}(z)$ is non-zero. Thus we have shown that $\hat{a}_0 \neq 0$. We can therefore scale $\hat{A}_{p-\hat{s}}(z)$ so that $\hat{a}_0 = 1$, and this scaled polynomial will still be a member of the null space of

$\mathbf{K}_{\hat{\gamma}}(p-\hat{s}, q-\hat{s})$. From now on, we assume that $\hat{a}_0 = 1$.

Clearly $\hat{\gamma}$, $\hat{A}_{p-\hat{s}}(z)$ are gain and pole polynomial solutions for the MEM $(p-\hat{s}, q-\hat{s})$ problem. Let us define $\hat{B}_{q-\hat{s}}(z)$ to be the corresponding zero polynomial with coefficients:

$$\hat{b}_n = \sum_{k=0}^n g[n-k] \hat{a}_k \quad (\text{B.4})$$

where we define $\hat{a}_k = 0$ for $k > p-\hat{s}$.

Now we prove that both $\hat{A}_{p-\hat{s}}(z)$ and $z^{-1}\hat{A}_{p-\hat{s}}(z)$ are members of the null space of $\mathbf{K}_{\hat{\gamma}}(p-\hat{s}+1, q-\hat{s}+1)$. Using (4.2):

$$\begin{aligned} & (\hat{a}_{p-\hat{s}}^* \cdots \hat{a}_0^* 0) \mathbf{K}_{\hat{\gamma}}(p-\hat{s}+1, q-\hat{s}+1) \begin{pmatrix} \hat{a}_{p-\hat{s}} \\ \vdots \\ \hat{a}_0 \\ 0 \end{pmatrix} \\ &= (\hat{a}_{p-\hat{s}}^* \cdots \hat{a}_0^*) \mathbf{K}_{\hat{\gamma}}(p-\hat{s}, q-\hat{s}) \begin{pmatrix} \hat{a}_{p-\hat{s}} \\ \vdots \\ \hat{a}_0 \end{pmatrix} \\ &= 0 \end{aligned} \quad (\text{B.5})$$

Since $\mathbf{K}_{\hat{\gamma}}(p-\hat{s}+1, q-\hat{s}+1) \geq 0$, this shows that $z^{-1}\hat{A}_{p-\hat{s}}(z)$ is a member of the null space of $\mathbf{K}_{\hat{\gamma}}(p-\hat{s}+1, q-\hat{s}+1)$.

Next, using (4.3) we can show that:

$$\begin{aligned}
& (0 \hat{a}_{p-\hat{s}}^* \cdots \hat{a}_0^*) \mathbf{K}_{\hat{\gamma}}(p-\hat{s}+1, q-\hat{s}+1) \begin{pmatrix} 0 \\ \hat{a}_{p-\hat{s}} \\ \vdots \\ \hat{a}_0 \end{pmatrix} & \text{(B.6)} \\
& = (\hat{a}_{p-\hat{s}}^* \cdots \hat{a}_0^*) \mathbf{K}_{\hat{\gamma}}(p-\hat{s}, q-\hat{s}) \begin{pmatrix} \hat{a}_{p-\hat{s}} \\ \vdots \\ \hat{a}_0 \end{pmatrix} - \hat{\gamma}^2 \left| \sum_{n=0}^{q-\hat{s}+1} g[q-\hat{s}+1-n] \hat{a}_n \right|^2 \\
& = -\hat{\gamma}^2 \left| \hat{b}_{q-\hat{s}+1} \right|^2 \\
& \leq 0
\end{aligned}$$

But since $\mathbf{K}_{\hat{\gamma}}(p-\hat{s}+1, q-\hat{s}+1) \geq 0$, this equation must actually equal zero. This shows that $\hat{A}_{p-\hat{s}}(z)$ is a member of the null space of $\mathbf{K}_{\hat{\gamma}}(p-\hat{s}+1, q-\hat{s}+1)$. In addition, since $\hat{\gamma}^2 = 1/\lambda_{\max} > 0$, we must have $\hat{b}_{q-\hat{s}+1} = 0$. Since both $\hat{A}_{p-\hat{s}}(z)$ and $z^{-1}\hat{A}_{p-\hat{s}}(z)$ are members of the null space of $\mathbf{K}_{\hat{\gamma}}(p-\hat{s}+1, q-\hat{s}+1)$, and these two polynomials correspond to linearly independent vectors of coefficients, the dimension of the null space of $\mathbf{K}_{\hat{\gamma}}(p-\hat{s}+1, q-\hat{s}+1)$ must be at least one greater than the dimension of the null space of $\mathbf{K}_{\hat{\gamma}}(p-\hat{s}, q-\hat{s})$.

Applying this argument recursively, we can show that for $s = 0, \dots, \hat{s}$, the polynomials $z^{-k}\hat{A}_{p-\hat{s}}(z)$ for $k = 0, \dots, s$ are all members of the null space of $\mathbf{K}_{\hat{\gamma}}(p-\hat{s}+s, q-\hat{s}+s)$. Since these members are all linearly independent, the dimension of the null space of $\mathbf{K}_{\hat{\gamma}}(p, q)$, which we assumed to be r , must be at least $\hat{s}+1$.

Thus $r \geq \hat{s} + 1$; however, we earlier showed that $r \leq \hat{s} + 1$. We conclude that $r = \hat{s} + 1$. Furthermore, the same recursive argument shows that $\hat{b}_n = 0$ for $n = q - r + 2, \dots, q$. This in turn implies the following:

- a) Matrix $\mathbf{K}_{\hat{\gamma}}(p - r + 1, q - r + 1) \geq 0$ with a null space with dimension one. There is a unique member of its null space, $\hat{A}_{p-r+1}(z)$, with leading coefficient of $\hat{a}_0 = 1$.
- b) For s in the range $0 \leq s < r$, the optimal gain for the *MEM* $(p - s, q - s)$ problem is exactly the same as the optimal gain $\hat{\gamma}^2 = 1/\lambda_{\max}$ for the *MEM* (p, q) problem. (This is because Lemma 1 guarantees that there is exactly one value of the gain for which the matrix $\mathbf{K}_{\hat{\gamma}}(p - s, q - s)$ is exactly positive semi-definite; $\hat{\gamma}$ is this gain value.)
- c) For s in the range $0 \leq s < r$, the null space of $\mathbf{K}_{\hat{\gamma}}(p - s, q - s)$ has dimension $r - s$, and is spanned by $\hat{A}_{p-r+1}(z), \dots, z^{-r+s+1}\hat{A}_{p-r+1}(z)$. (The dimension $r - s$ is a consequence of our proof that $\hat{s} = r - 1$. These vectors must be a basis because we have shown that all $r - s$ of them belong to the null space, and they are all linearly independent; thus they must span the null space.)
- d) For s in the range $r \leq s \leq p$, matrix $\mathbf{K}_{\hat{\gamma}}(p - s, q - s) > 0$. (This is true because $\mathbf{K}_{\hat{\gamma}}(p - s, q - s)$ is a principle minor of $\mathbf{K}_{\hat{\gamma}}(p, q)$, and thus must be at least positive semi-definite. It is strictly positive definite because $\hat{s} = r - 1$ is the largest value of s for which $\mathbf{K}_{\hat{\gamma}}(p - s, q - s)$ has a non-trivial null-space.)
- e) For s in the range $0 \leq s < r$, $A_{p-s}(z)$ is a solution to the *MEM* $(p - s, q - s)$ problem with $\hat{a}_0 = 1$, if and only if:

$$A_{p-s}(z) = \sum_{n=0}^{r-s-1} \phi_n z^{-n} \hat{A}_{p-r+1}(z) = \phi_{r-s-1}(z) \hat{A}_{p-r+1}(z) \quad (\text{B.7})$$

where $\phi_{r-s-1}(z)$ is some polynomial of order $r-s-1$ with leading coefficient $\phi_0 = 1$. (This is because $A_{p-s}(z)$ must be in the null space of $K_{\hat{\gamma}}(p-s, q-s)$, and therefore must be a linear combination of the basis vectors $\hat{A}_{p-r-1}(z), \dots, z^{-r+s+1} \hat{A}_{p-r-1}(z)$. Conversely, every such linear combination must be an element of the null space, and therefore must be a solution to the problem. We must have $\phi_0 = 1$ in order to make $\hat{a}_0 = 1$.)

Finally, for s in the range $0 \leq s < r$, let $A_{p-s}(z)$ be any pole polynomial solution to the $MEM(p-s, q-s)$ problem, so that $A_{p-s}(z) = \phi_{r-s-1}(z) \hat{A}_{p-r+1}(z)$ for some $r-s-1$ order polynomial $\phi_{r-s-1}(z)$ with $\phi_0 = 1$. Let $B_{q-s}(z)$ be the corresponding $q-s$ order zero polynomial with coefficients:

$$b_n = \sum_{k=0}^n g[n-k] a_k \quad \text{for } n=0, \dots, q-s \quad (\text{B.8})$$

where we define $a_k = 0$ for $k > p-s$. Substituting for a_k :

$$\begin{aligned} b_n &= \sum_{k=0}^n \sum_{l=0}^{r-s-1} g[n-k] \hat{a}_{k-l} \phi_l \\ &= \sum_{l=0}^{r-s-1} \phi_l \sum_{k=l}^n g[n-k] \hat{a}_{k-l} \\ &= \sum_{l=0}^{r-s-1} \phi_l \hat{b}_{n-l} \end{aligned} \quad (\text{B.9})$$

where \hat{a}_n and \hat{b}_n are the coefficients of the $\hat{A}_{p-r+1}(z)$ and $\hat{B}_{q-r+1}(z)$ polynomials.

But since $\hat{b}_n = 0$ for $n = q-r+2, \dots, q$, this implies that:

$$B_{q-s}(z) = \phi_{r-s-1}(z) \hat{B}_{q-r+1}(z) \quad (\text{B.10})$$

Thus $A_{p-s}(z)$ and $B_{q-s}(z)$ share up to $r-s-1$ common factors $\phi_{r-s-1}(z)$. The power spectrum, however, depends only on the ratio of $B_{q-s}(z)/A_{p-s}(z)$, and therefore is uniquely determined, and is identical to the *MEM* $(p-r+1, q-r+1)$ power spectrum $\hat{P}_{p-r+1, q-r+1}(z)$.

The multiplicity r must be less than or equal to the total number of eigenvalues, $r \leq p+1$. Also for $p+1 \geq n > q+1$, $K_{\gamma}(p-n+1, q-n+1) = R_{p-n+1} > 0$. Thus $r \leq q+1$. Combining gives $r \leq \min(p, q)+1$.

Appendix C - Proof of Theorem 3 and part A of Theorem 7

Assume that $R_p > 0$ and that λ_{\max} has multiplicity r . Set γ^2 in the range $0 \leq \gamma^2 \leq 1/\lambda_{\max}$, so that $K_\gamma(p, q) \geq 0$. For convenience, we will assume that if $\gamma^2 = \hat{\gamma}^2$ then $r=1$; the case $r > 1$ can be treated by replacing p and q everywhere below with $p-r+1$ and $q-r+1$ respectively. For notational convenience, let $K = K_\gamma(p, q)$. Let \hat{a}_p be a solution to the minimization problem in (5.1) for $n = p$:

$$\hat{a}_p \leftarrow \min_{a_p} a_p^H K a_p \quad (\text{C.1})$$

Then $\epsilon_p = \hat{a}_p^H K \hat{a}_p \geq 0$.

If $p = 0$, then $\hat{A}_p(z) = 1$, and the pole polynomial is stable. For $p > 1$, let us factor the polynomial $A_p(z) = (1 - \rho z^{-1})\psi(z)$ where $\psi(z)$ is a $p-1$ order polynomial with $\psi_0 = 1$, and ρ is one complex root of $A_p(z)$. Minimizing (B.3) over all a_1, \dots, a_p is equivalent to minimizing over all $\psi_1, \dots, \psi_{p-1}$ and ρ . In particular, if we factor the minimizing polynomial $\hat{A}_p(z) = (1 - \hat{\rho} z^{-1})\psi(z)$, and define vectors:

$$\underline{\psi}_+ = \begin{pmatrix} \psi_{p-1} \\ \vdots \\ \psi_0 \\ 0 \end{pmatrix} \quad \underline{\psi}_- = \begin{pmatrix} 0 \\ \psi_{p-1} \\ \vdots \\ \psi_0 \end{pmatrix} \quad (\text{C.2})$$

then

$$\hat{a}_p = \left(\underline{\psi}_+ \underline{\psi}_- \right) \begin{pmatrix} -\hat{\rho} \\ 1 \end{pmatrix} \quad (\text{C.3})$$

and $\hat{\rho}$ must be the solution to the following quadratic minimization problem:

$$\hat{\rho} = \min_{\rho} \begin{pmatrix} -\rho^* & 1 \end{pmatrix} \begin{pmatrix} \underline{\psi}_+^H \\ \underline{\psi}_-^H \end{pmatrix} \mathbf{K} \begin{pmatrix} \underline{\psi}_+ \\ \underline{\psi}_- \end{pmatrix} \begin{pmatrix} -\rho \\ 1 \end{pmatrix} \quad (\text{C.4})$$

This minimum occurs at:

$$\hat{\rho} = - \frac{\underline{\psi}_+^H \mathbf{K} \underline{\psi}_-}{\underline{\psi}_+^H \mathbf{K} \underline{\psi}_+} \quad (\text{C.5})$$

and the value at the minimum is:

$$\epsilon_p = \underline{\psi}_-^H \hat{\mathbf{K}} \underline{\psi}_- - \frac{\left(\underline{\psi}_+^H \hat{\mathbf{K}} \underline{\psi}_- \right) \left(\underline{\psi}_-^H \hat{\mathbf{K}} \underline{\psi}_+ \right)}{\underline{\psi}_+^H \hat{\mathbf{K}} \underline{\psi}_+} \quad (\text{C.6})$$

But $\epsilon_p \geq 0$ and by conjugate symmetry, $\underline{\psi}_+^H \mathbf{K} \underline{\psi}_- = \left(\underline{\psi}_-^H \mathbf{K} \underline{\psi}_+ \right)^*$. Thus combining (C.5) and (C.6):

$$\left| \hat{\rho} \right|^2 = \frac{\underline{\psi}_-^H \mathbf{K} \underline{\psi}_- - \epsilon_p}{\underline{\psi}_+^H \mathbf{K} \underline{\psi}_+} \quad (\text{C.7})$$

Using (C.2) and (4.2):

$$\underline{\psi}_+^H \mathbf{K} \underline{\psi}_+ = \begin{pmatrix} \psi_{p-1}^* & \cdots & \psi_0^* \end{pmatrix} \mathbf{K}_{\gamma(p-1, q-1)} \begin{pmatrix} \psi_{p-1} \\ \vdots \\ \psi_0 \end{pmatrix} > 0 \quad (\text{C.8})$$

where strict inequality must hold since $p > 1$, $r = 1$, $\psi_0 = 1$, and by Theorem 2, $\mathbf{K}_{\gamma(p-1, q-1)} > 0$. This guarantees that (C.5) is well-defined. Using (C.2) and (4.3):

$$\underline{\psi}_-^H \mathbf{K} \underline{\psi}_- = \begin{pmatrix} \psi_{p-1}^* & \cdots & \psi_0^* \end{pmatrix} \mathbf{K}_{\hat{\gamma}(p-1, q-1)} \begin{pmatrix} \psi_{p-1} \\ \vdots \\ \psi_0 \end{pmatrix} - \hat{\gamma}^2 |\hat{\mu}_q|^2 \quad (\text{C.9})$$

$$= \underline{\psi}_+^H \mathbf{K} \underline{\psi}_+ - \hat{\gamma}^2 |\hat{\mu}_q|^2$$

where we define:

$$\hat{\mu}_n = \sum_{k=0}^n g[n-k] \psi_k \quad (\text{C.10})$$

and where we define $\psi_k = 0$ for $k > p-1$. Combining (C.9) and (C.7):

$$|\hat{\rho}|^2 = 1 - \frac{\gamma^2 |\mu_q|^2 + \epsilon_p}{\underline{\psi}_+^H \mathbf{K} \underline{\psi}_+} \quad (\text{C.11})$$

The pole $\hat{\rho}$ must therefore be inside or on the unit circle. Since we chose the root arbitrarily, all poles in $\hat{A}_p(z)$ must be inside or on the unit circle.

If the gain γ^2 is below $\hat{\gamma}^2 = 1/\lambda_{\max}$, then $\mathbf{K} > 0$ and $\epsilon_p > 0$. Thus $|\hat{\rho}| < 1$, and every root of $\hat{A}_p(z)$ will be strictly inside the unit circle. (This proves part A of Theorem 7). If $\gamma^2 = \hat{\gamma}^2 = 1/\lambda_{\max}$, however, then $\epsilon_p = 0$, and $\hat{A}_p(z)$ will be the MEM(p, q) pole polynomial. All the roots of $\hat{A}_p(z)$ will be either inside or on the unit circle. From (C.11), a root of $\hat{A}_p(z)$ can be on the unit circle if and only if $\hat{\mu}_q = 0$. But then if we examine the zero polynomial corresponding to $\hat{A}_p(z)$:

$$\begin{aligned} \hat{b}_n &= \sum_{k=0}^n g[n-k] \hat{a}_k \\ &= \sum_{k=0}^n g[n-k] \psi_k - \rho g[n-k] \psi_{k-1} \\ &= \hat{\mu}_n - \rho \hat{\mu}_{n-1} \end{aligned} \quad (\text{C.12})$$

so that:

$$\hat{B}_q(z) = (1 - \rho z^{-1}) \hat{\mu}(z) \quad (\text{C.13})$$

where $\hat{\mu}(z)$ is a $q-1$ order polynomial with coefficients $\hat{\mu}_0, \dots, \hat{\mu}_{q-1}$. Thus if the $MEM(p, q)$ pole polynomial $\hat{A}_p(z)$ has a pole ρ on the unit circle, then the corresponding zero polynomial $\hat{B}_q(z)$ must have a canceling zero on the unit circle at ρ .

Now we will prove that all canceling pole-zero pairs in the $MEM(p, q)$ model must be on the unit circle. Suppose that $\hat{A}_p(z)$ and $\hat{B}_q(z)$ share a common factor $(1 - \hat{\rho}z^{-1})$. As above, $\hat{A}_p(z) = (1 - \hat{\rho}z^{-1}) \hat{\psi}(z)$, where $\hat{\psi}(z)$ is an order $p-1$ polynomial with leading coefficient $\hat{\psi}_0 = 1$. The coefficients of the corresponding zero polynomial must obey (C.12). This implies that:

$$\hat{B}_q(z) = (1 - \hat{\rho}z^{-1}) \left(\sum_{k=0}^{q-1} \hat{\mu}_k z^{-k} \right) + \hat{\mu}_q \quad (\text{C.14})$$

But we assumed that $(1 - \hat{\rho}z^{-1})$ was a factor of $\hat{B}_q(z)$; thus $\hat{\mu}_q = 0$. But then from (C.11), $|\hat{\rho}| = 1$, and so this canceling pole-zero pair must be on the unit circle.

Appendix D - Proof of Theorems 4 and 5

Proof of 4A)

Let $\hat{g}[n]$ be the causal impulse response of $\hat{B}_q(z)/\hat{A}_p(z)$. By definition, for $n < 0$, $\hat{g}[n] = 0$, and for $0 \leq n \leq q$:

$$\sum_{k=0}^n \hat{g}[n-k] \hat{a}_k = \hat{b}_k \quad (\text{D.1})$$

But the *MEM* (p, q) solution must also satisfy equation (3.13):

$$\sum_{k=0}^n g[n-k] \hat{a}_k = \hat{b}_k \quad \text{for } n = 0, \dots, q \quad (\text{D.2})$$

Define $\epsilon_g[n] = g[n] - \hat{g}[n]$, subtract (D.1) from (D.2), and write the result in matrix form:

$$\begin{pmatrix} 1 & & & 0 \\ \hat{a}_1 & 1 & & \\ \vdots & \cdot & \cdot & \\ \hat{a}_q & \cdot & \hat{a}_1 & 1 \end{pmatrix} \begin{pmatrix} \epsilon_g[0] \\ \vdots \\ \vdots \\ \epsilon_g[q] \end{pmatrix} = \mathbf{0} \quad (\text{D.3})$$

Since the matrix is invertible, $\epsilon_g[n] = 0$ for $n = 0, \dots, q$ and the first $q+1$ coefficients of the model impulse response must match the given data, $\hat{g}[n] = g[n]$.

Proof of 4B)

Let $\hat{P}_{p,q}(z)$ be the model power spectrum:

$$\hat{P}_{p,q}(z) = \hat{\gamma}^2 \frac{\hat{B}_q(z) \hat{B}_q^*(1/z^*)}{\hat{A}_p(z) \hat{A}_p^*(1/z^*)} \quad (\text{D.4})$$

and let $\hat{R}[n]$ be the correlations of the model. $\hat{R}[n]$ is just the inverse z-transform of $\hat{P}_{p,q}(z)$. Multiplying through by $\hat{A}_p(z)$, identifying the leading coefficients of $\hat{B}_q^*(1/z^*)/\hat{A}^*(1/z^*)$ as $\hat{g}^*[-n]$, and equating terms with equal powers of z , we get an equation similar in form to (3.16):

$$\hat{R}_p \hat{a} = \hat{\gamma}^2 \hat{G}_{p,q} \hat{b} \quad (\text{D.5})$$

where correlation matrix \hat{R}_p has $(n,m)^{\text{th}}$ element $\hat{R}[m-n]$, and $\hat{G}_{p,q}$ has $(n,m)^{\text{th}}$ element $\hat{g}[q-p+m-n]$. But the MEM(p,q) model $\hat{\gamma}$, \hat{a} , \hat{b} must also satisfy equation (3.16). Recognizing that $\hat{g}[n] = g[n]$ for $n \leq q$, and thus $\hat{G}_{p,q} = G_{p,q}$, subtracting (D.5) from (3.16) gives:

$$\begin{pmatrix} \epsilon_R[0] & \cdots & \epsilon_R[p] \\ \vdots & \ddots & \vdots \\ \epsilon_R[-p] & \cdots & \epsilon_R[0] \end{pmatrix} \begin{pmatrix} \hat{a}_p \\ \vdots \\ \hat{a}_1 \\ 1 \end{pmatrix} = \mathbf{0} \quad (\text{D.6})$$

where $\epsilon_R[n] = R[n] - \hat{R}[n]$ is the correlation matching error. To analyze these error values, let us extrapolate the tails of $\epsilon_R[n]$ by forward and backward prediction using the pole polynomial coefficients \hat{a}_k :

$$\epsilon_R[n] = -\hat{a}_1 \epsilon_R[n-1] - \cdots - \hat{a}_p \epsilon_R[n-p] \quad \text{for } n = p+1, p+2, \dots \quad (\text{D.7})$$

$$\epsilon_R[n] = -\hat{a}_1^* \epsilon_R[n+1] - \cdots - \hat{a}_p^* \epsilon_R[n+p] \quad \text{for } n = -p-1, -p-2, \dots$$

Using (D.6) and (D.7) we can show that:

$$\sum_{l=0}^p \sum_{k=0}^p \epsilon_R[n-k+l] \hat{a}_l^* \hat{a}_k = 0 \quad \text{for all } n \quad (\text{D.8})$$

If the polynomial $\hat{A}_p(z)$ has all its roots strictly inside the unit circle, then the extrapolated sequence $\epsilon_R[n]$ will decay exponentially to zero, and thus must be absolutely

summable. Therefore the Fourier transform $\epsilon_R(e^{j\omega})$ exists and is finite for all ω .

Fourier Transforming (D.8) gives:

$$\epsilon_R(e^{j\omega}) |\hat{A}_p(e^{j\omega})|^2 = 0 \quad (\text{D.9})$$

Since all of the roots of $\hat{A}_p(z)$ are strictly inside the unit circle, $|\hat{A}_p(e^{j\omega})|^2 > 0$ for all ω . Thus we must have $\epsilon_R(e^{j\omega}) = 0$ for all ω , which implies that $\epsilon_R[n] = 0$ for all n . Thus the model must match the given correlations.

Proof of 5A)

If $\hat{A}_p(z)$ has one or more roots ρ_i on the unit circle, $\rho_i = e^{j\omega_i}$, then the argument above breaks down. First of all, Theorem 3 guarantees that any poles on the unit circle must be canceled by matching zeroes. Thus the power spectrum of the model has a Region Of Convergence which includes the unit circle, and the correlations are still well-defined. The extrapolated tails of $\epsilon_R[n]$, however, may oscillate at the frequencies of the roots on the unit circle. Suppose that the roots of $\hat{A}_p(z)$ are all different. Then the tails of $\epsilon_R[n]$ will be bounded, but may not be absolutely summable. The Fourier Transform $\epsilon_R(e^{j\omega})$ may therefore contain impulses. In particular, since $|\hat{A}_p(e^{j\omega})|^2 = 0$ at the frequencies $\omega = \omega_i$, a non-zero solution for $\epsilon_R(e^{j\omega})$ in (D.9) will be a weighted sum of impulses located at the unit circle roots $e^{j\omega_i}$:

$$\epsilon_R(e^{j\omega}) = \sum_i \mu_i \delta(\omega - \omega_i) \quad (\text{D.10})$$

for some set of weights μ_i . Inverse Transforming:

$$\epsilon_R[n] = \sum_i \mu_i \rho_i^n \quad (\text{D.11})$$

More generally, if $\hat{A}_p(z)$ has roots ρ_i with multiplicities s_i , then $\epsilon_R(e^{j\omega})$ will be a weighted sum of impulses and impulse derivatives of order up to $2s_i - 1$. Since $\epsilon_R[n]$ is even, however, only even order derivatives can be present. Thus $\epsilon_R(e^{j\omega})$ will contain terms such as $\rho_i^n, n^2 \rho_i^n, \dots, n^{2s_i-2} \rho_i^n$.

Proof of 4C)

If $\hat{G}(z) = \hat{B}_q(z)/\hat{A}_p(z)$ is strictly minimum phase, then by Theorem 4A, its impulse response has the correct first $q + 1$ coefficients $g[0], \dots, g[q]$. Rewriting the recursive equation (3.10) to define the cepstra in terms of the minimum phase coefficients:

$$c[n] = \frac{1}{n} \sum_{k=0}^n k g[k] g[n-k] \quad (\text{D.12})$$

This equation provides a one-to-one mapping between the $g[0], \dots, g[q]$ and $c[1], \dots, c[q]$. Thus the first $q + 1$ cepstral coefficients $\hat{c}[n]$ must equal the specified values $c[1], \dots, c[q]$. Furthermore, the entropy of the model is $\log \hat{\gamma}^2$. By Theorem 1, however, this is the maximum possible entropy of any model matching the given data. Therefore, the *MEM* (p, q) solution is indeed the solution to the constrained Maximum Entropy problem. By Theorem 2, it is unique except for possible pole-zero cancellation if the multiplicity of λ_{\max} is greater than 1.

Proof of 5B)

Suppose that $\hat{B}_q(z)$ is not minimum phase, so that it has zeroes strictly outside the unit circle. Suppose there are q_0 zeroes outside the unit circle, $\rho_0, \dots, \rho_{q_0-1}$, and $q - q_0$ zeroes inside, $\rho_{q_0}, \dots, \rho_{q-1}$.

$$\hat{B}_q(z) = \prod_{i=0}^{q_0-1} (1 - \rho_i z^{-1}) \prod_{j=q_0}^{q-1} (1 - \rho_j z^{-1}) \quad (\text{D.13})$$

To calculate the model cepstral values $\hat{c}[1], \dots, \hat{c}[q]$, we need to factor the model power spectrum into a gain times a minimum phase factor times a maximum phase factor. In this case we get:

$$\hat{P}_{p,q}(z) = \gamma^2 \frac{D_q(z)}{\hat{A}_p(z)} \frac{D_q^*(1/z^*)}{\hat{A}_p^*(1/z^*)} \quad (\text{D.14})$$

where:

$$D_q(z) = \prod_{i=0}^{q_0-1} \left(1 - \frac{1}{\rho_i^*} z^{-1}\right) \prod_{j=q_0}^{q-1} (1 - \rho_j z^{-1}) \quad (\text{D.15})$$

$$\gamma^2 = \hat{\gamma}^2 \prod_{i=0}^{q_0-1} |\rho_i|^2$$

The impulse response $g'[n]$ of this minimum phase factor $D(z)/\hat{A}(z)$ satisfies:

$$\begin{pmatrix} 1 & & & 0 \\ \hat{a}_1 & \cdot & & \\ \vdots & \cdot & \cdot & \\ \hat{a}_q & \cdot & \hat{a}_1 & 1 \end{pmatrix} \begin{pmatrix} g'[0] \\ \vdots \\ \vdots \\ g'[q] \end{pmatrix} = \begin{pmatrix} d_0 \\ \vdots \\ \vdots \\ d_q \end{pmatrix} \quad (\text{D.16})$$

where the d_k are the coefficients of $D(z)$. Subtracting this from (D.2):

$$\begin{pmatrix} 1 & & & 0 \\ \hat{a}_1 & \cdot & & \\ \vdots & \cdot & \cdot & \\ \hat{a}_q & \cdot & \hat{a}_1 & 1 \end{pmatrix} \begin{pmatrix} g[0]-g'[0] \\ \vdots \\ \vdots \\ g[q]-g'[q] \end{pmatrix} = \begin{pmatrix} \hat{b}_0-d_0 \\ \vdots \\ \vdots \\ \hat{b}_q-d_q \end{pmatrix} \quad (\text{D.17})$$

Since the matrix on the left is invertible, and since the right hand side is non-zero, at least one of the coefficients of the minimum phase factor $g'[n]$ does not match $g[n]$. Since equation (D.12) defines a one-to-one mapping between the minimum phase coefficients $g'[0], \dots, g'[q]$ and the model cepstral values $\hat{c}[1], \dots, \hat{c}[q]$, the model cepstrum cannot match all the original cepstra $c[1], \dots, c[q]$.

Finally, the entropy will equal:

$$\begin{aligned} H &= \log \gamma^2 \\ &= \log \hat{\gamma}^2 + \sum_{i=0}^{q_0-1} \log |\rho_i|^2 \\ &> \log \hat{\gamma}^2 \end{aligned} \quad (\text{D.18})$$

where the strict inequality holds because $\rho_0, \dots, \rho_{q_0-1}$ are outside the unit circle, and thus have magnitudes greater than one. The *MEM* (p, q) model does not solve the original Maximum Entropy problem. We conclude that there does not exist any strictly positive, finite power spectrum $\hat{P}(ejw)$ meeting the Paley-Wiener condition which matches the given correlations and cepstra, and which achieves the maximum possible entropy.

Appendix E - Derivation of Fast Levinson-Style Algorithm

Assume that we already have the values a_{n-1} , d_{n-1} , f_{n-1} and M_{n-1} . Now using (4.2):

$$\mathbf{K}_\gamma(n, n+q-p) \begin{pmatrix} 0 \\ a_{n-1, n-1} \\ \vdots \\ a_{1, n-1} \\ 1 \end{pmatrix} = \begin{pmatrix} \phi_n \\ 0 \\ \vdots \\ 0 \\ \epsilon_{n-1} \end{pmatrix} - \gamma^2 \psi_n \begin{pmatrix} g^*[q-p] \\ \vdots \\ \vdots \\ \vdots \\ g^*[n+q-p] \end{pmatrix} \quad (\text{E.1})$$

$$\text{where } \phi_n = \sum_{k=0}^{n-1} \bar{R}(n-k) a_{k, n-1}$$

$$\psi_n = \sum_{k=0}^{n-1} g[q-p+n-k] a_{k, n-1}$$

Using (4.3):

$$\mathbf{K}_\gamma(n, n+q-p) \begin{pmatrix} d_{0, n-1} \\ \vdots \\ d_{n-1, n-1} \\ 0 \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ \vdots \\ 0 \\ v_n \end{pmatrix} \quad (\text{E.2})$$

$$\text{where } v_n = \sum_{k=0}^n \left[\mathbf{K}_\gamma(n, n+q-p) \right]_{n, k} d_{k, n-1}$$

$$\mathbf{K}_\gamma(n, n+q-p) \begin{pmatrix} f_{0, n-1} \\ \vdots \\ f_{n-1, n-1} \\ 0 \end{pmatrix} = \begin{pmatrix} g^*[q-p] \\ \vdots \\ \vdots \\ g^*[n+q-p] \end{pmatrix} + \begin{pmatrix} 0 \\ \vdots \\ 0 \\ \mu_n \end{pmatrix} \quad (\text{E.3})$$

$$\text{where } \mu_n = \sum_{k=0}^{n-1} \left[\mathbf{K}_\gamma(n, n+q-p) \right]_{n, k} f_{k, n-1} - g^*[n+q-p]$$

We can now compute a_n by combining linear multiples of d_{n-1} and f_{n-1} , then compute d_n by combining linear multiples of d_{n-1} and a_n , and then compute f_n by combining linear multiples of f_{n-1} and a_n . The value of ϵ_n can be found by direct substi-

tution. These update formulas are given in the algorithm listing in section 5.

To compute M_n recursively, we start by expanding d_n and f_n in terms of d_{n-1} , f_{n-1} and a_n :

$$M_n = \left[\begin{pmatrix} d_{n-1}^H & 0 \\ f_{n-1}^H & 0 \end{pmatrix} - \frac{1}{\epsilon_n} \begin{pmatrix} v_n^* a_n^H \\ \mu_n^* a_n^H \end{pmatrix} \right] K_\gamma(n, n+q-p) \cdot \quad (E.4)$$

$$\left[\begin{pmatrix} d_{n-1} & f_{n-1} \\ 0 & 0 \end{pmatrix} - \frac{1}{\epsilon_n} \begin{pmatrix} v_n a_n & \mu_n a_n \end{pmatrix} \right]$$

Writing this out as a sum of four terms, and using (4.2) and (5.3), gives the update formula for M_n in section 5. Finally, we can avoid having to calculate v_n and μ_n directly by noting that:

$$\begin{pmatrix} d_{n-1}^H & 0 \\ f_{n-1}^H & 0 \end{pmatrix} K_\gamma(n, n+q-p) a_n^H = \begin{pmatrix} d_{n-1}^H & 0 \\ f_{n-1}^H & 0 \end{pmatrix} \begin{pmatrix} 0 \\ \vdots \\ 0 \\ \epsilon_n \end{pmatrix} \quad (E.5)$$

$$= \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

But also:

$$\begin{pmatrix} d_{n-1}^H & 0 \\ f_{n-1}^H & 0 \end{pmatrix} K_\gamma(n, n+q-p) a_n^H \quad (E.6)$$

$$= \begin{pmatrix} d_{n-1}^H & 0 \\ f_{n-1}^H & 0 \end{pmatrix} K_\gamma(n, n+q-p) \left[\begin{pmatrix} 0 \\ a_{n-1} \end{pmatrix} + \begin{pmatrix} d_{n-1} & f_{n-1} \\ 0 & 0 \end{pmatrix} \begin{pmatrix} -\phi_n \\ \gamma^2 \psi_n \end{pmatrix} \right]$$

$$\begin{aligned}
&= \begin{pmatrix} 1 & 0 & \cdots & 0 & \nu_n^* \\ g[q-p] & \cdot & \cdot & \cdot & g[n+q-p] + \mu_n^* \end{pmatrix} \begin{pmatrix} 0 \\ \underline{a}_{n-1} \end{pmatrix} + \mathbf{M}_{n-1} \begin{pmatrix} -\phi_n \\ \gamma^2 \psi_n \end{pmatrix} \\
&= \begin{pmatrix} \nu_n^* \\ \psi_n + \mu_n^* \end{pmatrix} + \mathbf{M}_{n-1} \begin{pmatrix} -\phi_n \\ \gamma^2 \psi_n \end{pmatrix}
\end{aligned}$$

Equating these gives the formula relating ν_n , μ_n to ϕ_n , ψ_n given in section 5. Thus we can derive ν_n and μ_n from this formula, rather than calculating them directly. This is particularly advantageous, since we no longer need to calculate the actual elements of the $\mathbf{K}_\gamma(n, n+q-p)$ matrix.

Appendix F - Proof of Theorem 6

The proof of this theorem relies on two key formulas involving the ϵ_n . The first formula is:

$$A^H K_\gamma(p, q) A = \begin{pmatrix} \epsilon_0 & 0 \\ \cdot & \cdot \\ 0 & \epsilon_p \end{pmatrix} \quad (\text{F.1})$$

$$\text{where } A = \begin{pmatrix} 1 & a_{1,1} & \cdot & a_{p,p} \\ & 1 & & \cdot \\ & & \cdot & a_{1,p} \\ 0 & & & 1 \end{pmatrix}$$

To prove this formula, note that for $n \leq m$, the $(n, m)^{th}$ element of the matrix on the left is:

$$\begin{aligned} \left(a_n^H Q^H \right) K_\gamma(p, q) \begin{pmatrix} a_m \\ Q \end{pmatrix} &= \left(a_n^H Q^H \right) K_\gamma(m, m+q-p) a_m & (\text{F.2}) \\ &= \left(a_n^H Q^H \right) \begin{pmatrix} 0 \\ \vdots \\ 0 \\ \epsilon_m \end{pmatrix} \\ &= \begin{cases} \epsilon_n & \text{if } n = m \\ 0 & \text{if } n < m \end{cases} \end{aligned}$$

where we used (4.2) in the first line and (5.3) in the second. The case $n > m$ can be treated similarly. We now invoke Sylvester's law of inertia[18,19] which states that since A is invertible, the matrix $A^H K_\gamma(p, q) A$ must have the same number of positive eigenvalues as $K_\gamma(p, q)$, the same number of negative eigenvalues, and the same

number of zero eigenvalues. The eigenvalues of $A^H K_\gamma(p, q) A$, however, are $\epsilon_0, \dots, \epsilon_p$. Thus, using Lemma 1, if $0 \leq \gamma^2 < 1/\lambda_{\max}$, then $K_\gamma(p, q) > 0$ and all the values of ϵ_n must be strictly positive. If $\gamma^2 > 1/\lambda_{\max}$, then $K_\gamma(p, q)$ has at least one negative eigenvalue, so at least one value ϵ_n is negative. If $\gamma^2 = 1/\lambda_{\max}$, then $K_\gamma(p, q) \geq 0$, it will have r eigenvalues equal to zero, and thus r of the values ϵ_n will be zero and the rest will be strictly positive.

To prove that the sequence of ϵ_n is decreasing for $K_\gamma(p, q) \geq 0$, we combine the formulas for ϵ_n and ν_n, μ_n in the Mullis-Roberts algorithm in section 5:

$$\begin{aligned} \epsilon_n &= \epsilon_{n-1} - \begin{pmatrix} \nu_n & \mu_n \end{pmatrix} \begin{pmatrix} \phi_n \\ -\gamma^2 \psi_n \end{pmatrix} \\ &= \epsilon_{n-1} - \begin{pmatrix} \phi_n^* & -\gamma^2 \psi_n^* \end{pmatrix} M_{n-1} \begin{pmatrix} \phi_n \\ -\gamma^2 \psi_n \end{pmatrix} - \gamma^2 |\psi_n|^2 \end{aligned} \quad (\text{F.3})$$

Now if $K_\gamma(p, q) \geq 0$, then by (4.2) $K_\gamma(n-1, n-1+q-p) \geq 0$, and by (5.6), $M_{n-1} \geq 0$ also. Thus $\epsilon_n \leq \epsilon_{n-1}$ for all n . For $\gamma^2 \leq 1/\lambda_{\max}$, therefore, the values of ϵ_n decrease with n , and are non-negative. For $\gamma^2 = 1/\lambda_{\max}$, the last r values must be zero.

For $K_\gamma(p, q) > 0$, all the ϵ_n will be strictly positive, and therefore we can run the Mullis-Roberts algorithm for steps $n = 0, \dots, p$ to calculate all the \mathbf{a}_n, ϵ_n . If $K_\gamma(p, q) \geq 0$ with an r dimensional null space, then $\epsilon_n > 0$ for $n = 0, \dots, p-r$ but $\epsilon_{p-r+1} = 0$. Thus we must stop the Mullis-Roberts algorithm at step $n = p-r+1$. However, Theorem 2 guarantees that the vectors $\mathbf{a}_n = (0 \mathbf{a}_{n-1}^T)^T$ for

$n = p - r + 2, \dots, p$ will be elements of the null space of $K_\gamma(n, n + q - p)$, and thus satisfy the assumptions of this theorem with $\epsilon_n = 0$ for $n = p - r + 2, \dots, p$.

Appendix G - Proof of Theorem 7

Let $0 \leq \gamma^2 < 1/\lambda_{\max}$. By Theorem 6, the Mullis-Roberts algorithm will be able to calculate the unique p^{th} order solution \mathbf{a}_p, ϵ_p .

Appendix C proves that for $\gamma^2 < \hat{\gamma}^2$, the pole polynomial solution to (5.1) is strictly stable with all poles strictly inside the unit circle. To prove part B of Theorem 7, let us define the zero polynomial $B_q(z)$ by substituting the coefficients of $A_p(z)$ into the recursion (3.13). Following the proof in Appendix D, we can show that the leading coefficients of the impulse response of our model exactly match the given values $g[0], \dots, g[q]$. The correlation matching proof proceeds similarly to that in Appendix D, except that we find that:

$$\begin{pmatrix} \epsilon_R[0] & \cdots & \epsilon_R[p] \\ \vdots & \ddots & \vdots \\ \epsilon_R[-p] & \cdots & \epsilon_R[0] \end{pmatrix} \begin{pmatrix} \hat{a}_p \\ \vdots \\ \hat{a}_1 \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ \epsilon_p \end{pmatrix} \quad (\text{G.1})$$

We can extrapolate $\epsilon_R[n]$ forwards and backwards as in Appendix D. We can then show that:

$$\sum_{l=0}^p \sum_{k=0}^p \epsilon_R[n-k+l] \hat{a}_l^* \hat{a}_k = \epsilon_p \delta[n] \quad \text{for all } n \quad (\text{G.2})$$

Since the pole polynomial is guaranteed by part A to be strictly stable, the extrapolated tails decay exponentially and are thus absolutely summable. We can thus Fourier Transform both sides of (G.2) to give:

$$\epsilon_R(e^{j\omega}) \left| A_p(e^{j\omega}) \right|^2 = \epsilon_p \quad (\text{G.3})$$

Dividing through by $|A_p(e^{j\omega})|^2$ and inverse transforming:

$$R[n] - \hat{R}[n] = \int_{-\pi}^{\pi} \frac{\epsilon_p}{|A_p(e^{j\omega})|^2} e^{j\omega n} \frac{d\omega}{2\pi} \quad (\text{G.4})$$

Appendix H - High Order Minimum Phase AR Approximations to Non-Minimum Phase Models

To find a minimum phase all-pole (AR) model $\epsilon_N/A_N(z)A_N^*(1/z^*)$ which matches $R[0]$ and $c[1]$ and which achieves maximum entropy for a given model order, we will try searching over the reflection coefficient domain. The reflection coefficients k_1, \dots, k_N are related to an N^{th} order polynomial $A_N(z)$ by a Levinson recursion, which builds an n^{th} order polynomial $A_n(z)$ from linear combinations of the $n-1^{\text{th}}$ order polynomial $A_{n-1}(z)$:

$$A_n(z) = A_{n-1}(z) + k_n z^{-n} A_{n-1}^*(1/z^*) \quad \text{for } n = 1, \dots, N \quad (\text{H.1})$$

where $A_0(z) \equiv 1$. Simply by keeping all k_n less than 1 in magnitude, we can force the all-pole model to be minimum phase. Furthermore, if we choose the model gain ϵ_N by the following recursive calculation:

$$\epsilon_0 = R[0] \quad (\text{H.2})$$

$$\epsilon_n = \epsilon_{n-1} (1 - |k_n|^2)$$

then each n^{th} order model $\epsilon_n/A_n(z)A_n^*(1/z^*)$ will have the zeroth correlation coefficient equal to $R[0]$. The model entropy will be:

$$H = \log \epsilon_N = \log R[0] + \sum_{n=1}^N \log (1 - |k_n|^2) \quad (\text{H.3})$$

Furthermore, the first cepstral coefficient will be

$$c[1] = -a_{1,N} = -\left(k_1 + \sum_{n=2}^N k_n k_{n-1}^*\right) \quad (\text{H.4})$$

The Maximum Entropy N^{th} order all-pole model matching $R[0]$ and $c[1]$ can now be found by maximizing (H.3) subject to the constraint (H.4) and subject to $|k_n| < 1$.

Introducing Lagrange multiplier λ for constraint (H.4), building the Lagrangian, and then differentiating to find the critical point, we can show that the reflection coefficients achieving maximum entropy while meeting the constraints must satisfy: (we assume all quantities are real-valued)

$$\frac{2k_i}{1 - |k_i|^2} = \lambda (k_{i-1} + k_{i+1}) \quad (\text{H.5})$$

where we define $k_0 = 1$ and $k_{N+1} = 0$. An exact AR(∞) solution for $c[1] = b_1 = -1$ is $\lambda = 1$ and $k_n = 1/(n+1)$. For finite order AR models with $c[1]$ somewhat more negative than -1, the solution will generally have λ slightly greater than 1, and k_n slightly greater than $1/(n+1)$. A relaxation algorithm based on equation (H.5) can be used to find these values. Choose an appropriate AR model order N , try an initial guess of $k_n = 1/(n+1)$, and guess an initial value for λ . Now repeatedly solve (H.5) for k_n for $n=1, \dots, N$. Iterate until the estimates converge. While iterating, compute the value of $c[1]$ in (H.4). If this is larger (smaller) than desired, then decrease (increase) λ .

Appendix I - Asymptotic Behavior of Correlations, Cepstral Estimates

To estimate the asymptotic behavior of the correlations and cepstral estimates derived from a periodogram, let us approximate the complex-valued, zero-mean Gaussian sequence $x[n]$ as periodic with period N . As $N \rightarrow \infty$, we would expect this approximation to give asymptotically valid formulas for the means and variances of estimators based on the data. (See, for example [25] or [26]). The power spectrum will be a line spectrum:

$$P_x(\omega) = \sum_{k=0}^{N-1} P_x(\omega_k) \delta(\omega - \omega_k) \quad \text{for } \omega_k = \frac{2\pi k}{N} \quad (\text{I.1})$$

and the periodic correlations and cepstra are defined by:

$$\hat{R}[n] = \frac{1}{N} \sum_{k=0}^{N-1} P_x(\omega_k) e^{j\omega_k n} \quad (\text{I.2})$$

$$\hat{c}[n] = \frac{1}{N} \sum_{k=0}^{N-1} \log P_x(\omega_k) e^{j\omega_k n}$$

Let $X(\omega_k)$ be the Discrete Fourier Transform (DFT) of $x[n]$:

$$X(\omega_k) = \sum_{n=0}^{N-1} x[n] e^{j\omega_k n} \quad (\text{I.3})$$

It is easy to show that the real and imaginary parts of $\frac{1}{\sqrt{N}}X(\omega_k)$ are independent

Gaussian random variables with zero mean and variance $\frac{1}{2}P_x(\omega_k)$:

$$\text{E} \left[\frac{1}{\sqrt{N}} \text{Re} X(\omega_k) \right] = \text{E} \left[\frac{1}{\sqrt{N}} \text{Im} X(\omega_k) \right] = 0 \quad (\text{I.4})$$

$$\text{E} \left[\frac{1}{N} \text{Re} X(\omega_k) \text{Re} X(\omega_l) \right] = \text{E} \left[\frac{1}{N} \text{Im} X(\omega_k) \text{Im} X(\omega_l) \right]$$

$$= \begin{cases} \frac{1}{2} P_x(\omega_k) & \text{for } k=0 \\ 0 & \text{else} \end{cases} \quad (\text{I.5})$$

$$E \left[\frac{1}{N} \text{Re} X(\omega_k) \text{Im} X(\omega_l) \right] = 0$$

Given $x[n]$, let us estimate the power as equal to the periodogram, $\hat{P}_x(\omega_k) = \frac{1}{N} |X(\omega_k)|^2$. Inverse transform the periodogram and the log periodogram to get the correlation and cepstral estimates $\hat{R}[n]$ and $\hat{c}[n]$ respectively. To compute the mean and variance of the estimates, we will use the formula:

$$\text{Cov}[u, v] = E[uv] - E[u]E[v] \quad (\text{I.6})$$

and where x, y, z and w are Gaussian:

$$E[xyzw] = E[xy]E[zw] + E[xz]E[yw] + E[xw]E[yz]$$

Then:

$$E \left[\hat{R}[n] \right] = \frac{1}{N} \sum_{k=0}^{N-1} E \left[\frac{1}{N} |X(\omega_k)|^2 \right] e^{j\omega_k n} = R[n] \quad (\text{I.7})$$

$$\begin{aligned} \text{Cov} \left[\hat{R}[n], \hat{R}[m] \right] &= \frac{1}{N^2} \sum_{k=0}^{N-1} \sum_{l=0}^{N-1} \text{Cov} \left[\frac{1}{N} |X(\omega_k)|^2, \frac{1}{N} |X(\omega_l)|^2 \right] e^{j(\omega_k n - \omega_l m)} \\ &= \frac{1}{N^2} \sum_{k=0}^{N-1} P_x^2(\omega_k) e^{j\omega_k(n-m)} \end{aligned} \quad (\text{I.8})$$

To compute the statistics of the cepstral estimates, we will use formulas from [27]:

$$\begin{aligned} E \left[\hat{c}[n] \right] &= \frac{1}{N} \sum_{k=0}^{N-1} E \left[\log \frac{1}{N} |X(\omega_k)|^2 \right] e^{j\omega_k n} \\ &= \frac{1}{N} \sum_{k=0}^{N-1} \left(\log P_x(\omega_k) + \int_0^{\infty} e^{-t} \log t \, dt \right) e^{j\omega_k n} \end{aligned}$$

$$= c[n] - \gamma \delta[n] \quad (\text{I.9})$$

$$\begin{aligned} \text{Cov} \left[\hat{c}[n], \hat{c}[m] \right] &= \frac{1}{N^2} \sum_{k=0}^{N-1} \sum_{l=0}^{N-1} \text{Cov} \left[\log \frac{1}{N} |X(\omega_k)|^2, \log \frac{1}{N} |X(\omega_l)|^2 \right] e^{j(\omega_k n - \omega_l m)} \\ &= \frac{1}{N^2} \sum_{k=0}^{N-1} \left(\int_0^{\infty} e^{-t} (\log t)^2 dt - \gamma^2 \right) e^{j\omega_k(n-m)} \\ &= \frac{1}{N} \frac{\pi^2}{6} \delta(n-m) \end{aligned} \quad (\text{I.10})$$

where $\gamma \approx .577$ is Euler's constant.

References

- [1] John P. Burg, *Maximum Entropy Spectral Analysis*, Ph.D. Thesis, Stanford University, May 1975.
 - [2] John P. Burg, "Maximum Entropy Spectral Analysis," *Proc. 37th Meeting of the Soc. Expl. Geophysicists*, reprinted in *Modern Spectrum Analysis*, ed. Donald G. Childers, IEEE Press 1978, New York, 1967.
 - [3] E.T. Jaynes, "New Engineering Applications of Information Theory" in *Proc. First Symposium on Engineering Applications of Random Function Theory and Probability*, ed. J.L. Bogdanoff, F. Kozin, New York: John Wiley, 1963, pp. 163-203.
 - [4] E.T. Jaynes, "Prior Probabilities," *IEEE Trans. Syst. Sci. Cybern.*, vol. SSC-4, pp. 227-241, 1968.
 - [5] J. Makhoul, "Linear Prediction: A Tutorial Review," *Proc. IEEE*, vol. 63, no. 4, pp. 561-580, reprinted in *Modern Spectrum Analysis*, ed. Donald G. Childers, IEEE Press, New York 1978, April 1975.
 - [6] Donald G. Childers, *Modern Spectrum Analysis*, New York: IEEE Press, 1978.
 - [7] Steven M. Kay and Stanley Lawrence Marple, "Spectrum Analysis - A Modern Perspective," *Proc. IEEE*, vol. 69, no. 11, pp. 1380-1418, Nov 1981.
 - [8] Miguel A. Lagunas-Hernandez, M. Eugenia Santamaria-Perez, and Anibal R. Figueiras-Vidal, "ARMA Model Maximum Entropy Power Spectral Estimation," *IEEE Trans. Acoustics, Speech, Sig. Proc.*, vol. ASSP-32, no. 5, pp. 984-990, Oct 1984.
 - [9] Clifford T. Mullis and Richard A. Roberts, "The Use of Second-Order Information in the Approximation of Discrete-Time Linear Systems," *IEEE Trans. Acoustics, Speech, and Signal Processing*, vol. ASSP-24, no. 3, pp. 226-238, June 1976.
 - [10] V.F. Pisarenko, "The Retrieval of Harmonics from a Covariance Function," *Geophys. J. Royal Astr. Soc.*, vol. 33, pp. 347-366, 1973.
 - [11] R.B. Blackman and J.W. Tukey, *The Measurement of Power Spectra*, New York: Dover, 1958.
 - [12] Gary E. Kopec, A.V. Oppenheim, and J.M. Tribolet, "Speech Analysis by Homomorphic Prediction," *IEEE Trans. Acoustics, Speech, Sig. Proc.*, vol. ASSP-25, no. 1, pp. 40-49, Feb 1977.
 - [13] B. Yegnanarayana, "Speech Analysis by Pole-Zero Decomposition of Short Time Spectra," *Signal Processing*, vol. ???, pp. 5-17.
 - [14] David Smith, *Variational Methods in Optimization*, Englewood Cliffs, NJ: Prentice Hall, Inc., 1974.
 - [15] James A. Cadzow, "High Resolution Spectral Estimation: An Assessment," *Trends and Perspectives in Signal Processing*, pp. 16-19, June 1983.
-

- [16] Norbert Wiener, *Extrapolation, Interpolation and Smoothing of Stationary Time Series*, Cambridge, MA: MIT Press, 1949.
 - [17] Alan V. Oppenheim and Ronald W. Schaffer, *Digital Signal Processing*, Englewood Cliffs, NJ: Prentice-Hall, 1975.
 - [18] Gilbert Strang, *Linear Algebra and its Applications*, New York: Academic Press, 1976.
 - [19] F.R. Gantmacher, *The Theory of Matrices*, New York, N.Y.: Chelsea Publishing Co., 1959.
 - [20] Bruce R. Musicus, "Levinson and Fast Choleski Algorithms for Toeplitz and Almost Toeplitz Matrices," (*to be submitted*).
 - [21] B. Friedlander, M. Morf, T. Kailath, and L. Ljung, "New Inversion Formulas for Matrices Classified In Terms of Their Distance from Toeplitz Matrices," *Lin. Alg. and Appl.*, vol. 27, pp. 31-60, 1979.
 - [22] B.S. Garbow, J.M. Boyle, J.J. Dongarra, and C.B. Moler, *Matrix Eigensystem Routines - EISPACK Guide Extension*, New York: Springer-Verlag, 1977.
 - [23] J. Durbin, "Efficient Estimation of Parameters in Moving Average Models," *Biometrika*, vol. 46, pp. 306-316, 1959.
 - [24] Louis L. Scharf and James C. Luby, "Statistical Design of Autoregressive-Moving Average Digital Filters," *IEEE Acoust., Speech, Sig. Proc.*, vol. ASSP-27, no. 3, pp. 240-247, June 1979.
 - [25] Ulf Grenander and Gabor Szego, *Toeplitz Forms and Their Applications*, Berkley: Univ. of Calif. Press, 1958.
 - [26] Robert Molten Gray, "On the Asymptotic Eigenvalue Distribution of Toeplitz Matrices," *IEEE Trans. Info. Theory*, vol. IT-18, no. 6, pp. 725-730, Nov. 1972.
 - [27] I.S. Gradshteyn and I.M. Ryzhik, *Table of Integrals, Series and Products*, New York: Academic Press, 1965.
-
