Mathematics Notes

Note 85

August 1985

Preprocessing Techniques for the Estimation of SEM Parameters from Multiple Measurements

Sung-won Park
J. T. Cordaro
University of New Mexico
Albuquerque NM 87131

Abstract

Two kinds of pole estimators for given noisy transient data are presented. One is a SVD-based method and the other is the iterative preprocessing algorithm (IPA). The SVD method estimates the characteristic equation coefficients from the weakest eigenvectors when the system order is overdetermined. The IPA is shown to converge to the maximum-likelihood estimator for some range of SNR. It not only reduces the computational burden of the currently existing iterative algorithms but also improves stability. The approximate IPA (AIPA), which further reduces the computational burden, is described. The Cramer-Rao lower bound for the system characteristic equation coefficients is calculated.

The IPA is extended to pole estimation for given multiple data sets. This technique is applied to estimating the natural frequencies of an electromagnetic scatterer. An iterative scheme is described for estimating other parameters of the singularity expansion method (SEM) formalism.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

In a variety of applications it is desirable to determine the natural frequencies or poles of a system from a noisy observation of its impulse response. A direct approach of minimizing an error with respect to poles and residues is a highly nonlinear regression problem. Except for low order cases this nonlinear regression problem is difficult to solve directly.

In general, it is easier to estimate the system characteristic equation first, and then find its roots from the estimated equation. Let $\hat{x}_{ML}$ be the maximum likelihood estimate of x. The invariance property of the maximum-likelihood estimator [1] states that if f(x) is an invertible function defined for all x, then the maximum likelihood estimate of f is just $f(\hat{x}_{ML})$. In our problem if the estimate of the characteristic equation coefficients is of maximum likelihood, then the roots of that equation are the maximum likelihood estimate of the system z-poles.

The pole estimation techniques are extended to the SEM (singularity expansion method) parameter estimation. The SEM formalism began from experimental observations concerning the transient electromagnetic response of complicated scatterers such as missiles and aircraft [2,3,4]. Instead of analyzing various parameters of a certain scatterer analytically, we can compute SEM parameters directly from the induced transient current data. It was observed that damped sinusoids were dominant features of typical transient responses [2]. The response of a scatterer to an incident impulse plane electromagnetic wave is expressed

as a sum of complex exponentials. This sum depends on a few parameters, namely natural frequencies which do not depend on an observation location and an incidence direction, coupling coefficients which describe the coupling between the incident field and the scatterer, and natural modes which describe the spatial amplitude variation in the induced current [2,3,4].

In Chapter 2, we review the covariance method and the singular value decomposition (SVD) method. Also a new algorithm is described to extract the reduced characteristic equation from the weakest eigenvectors when the system order is overdetermined.

In Chapter 3, the iterative preprocessing algorithm (IPA) is presented. This is related to the Steiglitz-McBride algorithm [5,6]. But this IPA not only reduces significantly the computational burden of the existing algorithm; it also improves the stability. The approximate iterative preprocessing algorithm (AIPA), which further reduces the computational burden, is discussed. The AIPA for the pure sinusoid case is related to Kay's iterative filtering algorithm (IFA) [7].

In Chapter 4, the Cramer-Rao bound for the system transfer function parameters is described. This is a very important tool for evaluating different estimators.

Some simulation results for the new SVD method, the IPA, and the AIPA are given in Chapter 5. The sample variances of the coefficients using the IPA lie exactly on the C-R bound curve for moderate SNR. This argues that the IPA converges to the maximum likelihood estimator for some range of SNR values.

In Chapter 6, the general SEM formalism is reviewed and a new algorithm is described for estimating natural frequencies given multiple

observations. Also the C-R bound is calculated for the characteristic equation coefficients for given multiple data sets.

In Chapter 7, an iterative scheme is proposed for estimating SEM parameters, such as coupling coefficients, natural modes, and normalization factors.

Some experimental results for SEM parameter estimation are given in Chapter 8.

CHAPTER 2

SVD METHOD

2.1 COVARIANCE METHOD AND SVD METHOD

Suppose that there are N samples of the observed data sequence $y(nT)$ consisting of K complex exponentials plus a zero-mean, uncorrelated noise $e(nT)$ such that

$$y(nT) = \sum_{i=1}^{K} c_i \exp(s_i nT) + e(nT) \quad \text{for } n = 0, 1, \ldots, N-1 \tag{1}$$

where (s_i) are the poles, (c_i) are the residues, and T is the time increment between successive samples. Let

$$w_n = \sum_{i=1}^{K} c_i z_i^{\,n} . \tag{2}$$

where the $z_i = \exp(s_i T)$ are the poles in terms of the z-transform variable. Then with no loss of generality, (1) becomes

$$y_n = w_n + e_n \tag{3}$$

The characteristic equation with respect to T is

$$\hat{B}(z) = b_0 z^K + b_1 z^{K-1} + \cdots + b_K \tag{4}$$

Once we obtain an estimate of $\hat{B}(z)$ from the data sequence (y_n), finding the roots of $\hat{B}(z) = 0$ is an algebraic procedure. The inverse mapping

$$s_i = [\log|z_i| + j\{\arg(z_i)\}]/T$$

is used to convert z-poles to s-poles, and residues can be found by a linear least squares method. Thus the main problem is estimating the coefficients (b_i).

In terms of the difference equation

$$\sum_{i=0}^{K} b_i w_{n-i} = 0 \quad \text{for } n = K, K+1, \ldots, N-1 \tag{5}$$

Define $d_n = \sum_{i=0}^{K} b_i e_{n-i}$ for $n = K, K+1, \ldots, N-1$.

With noise contaminated data, (5) becomes

$$\sum_{i=0}^{K} b_i y_{n-i} = d_n \quad \text{for } n = K, K+1, \ldots, N-1 \tag{6}$$

where the (d_n) are due to noise and are termed the equation error. In matrix form (6) becomes

$$\begin{bmatrix} y_0 & y_1 & \cdots & y_K \\ y_1 & y_2 & \cdots & y_{K+1} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ y_{N-K-1} & y_{N-K} & \cdots & y_{N-1} \end{bmatrix} \begin{bmatrix} b_K \\ b_{K-1} \\ \cdot \\ \cdot \\ b_0 \end{bmatrix} = \begin{bmatrix} d_K \\ d_{K+1} \\ \cdot \\ \cdot \\ \cdot \\ d_{N-1} \end{bmatrix}$$

$$Y \qquad x \qquad = \qquad d \tag{7}$$

By letting $b_0 = 1$, (7) becomes

$$Y_L x_I + y_R = d \tag{8}$$

where Y_L is the matrix consisting of the first K columns of Y, y_R is the (K+1)th column of Y, and $x_I = [b_K, b_{K-1}, \ldots, b_1]^T$. To minimize d^*d of (7), we need to solve

$$Y_L^* Y_L x_I = - Y_L^* y_R \tag{9}$$

This approach to estimating the (b_i) is called the covariance method [8]. It has been observed that this method has a statistical bias in the estimates [9]. The least squares process converges as N becomes large, but it yields biased parameter values which are a function of the standard deviation of the noise. It is unfortunate that even for small noise levels, this bias will produce large errors on the estimate s_i. Kay [10] examined a similar problem in autoregressive spectral estimation.

The singular value decomposition (SVD) approach deals directly with (7). The SVD is used to solve the homogeneous equation, i.e. to find the null space of the Y matrix. The SVD of Y is

$$Y = USV^* \tag{10}$$

where $S = \begin{bmatrix} \sigma_1 & & & 0 \\ & \sigma_2 & & \\ & & \ddots & \\ 0 & & & \sigma_{K+1} \\ \hline & & 0 & \end{bmatrix}$ } N-2K-1

Here (σ_i) are termed the singular values, which are the square roots of the eigenvalues of Y^*Y, and $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_{K+1}$; U and V are unitary matrices of orders (N-K)x(N-K) and (K+1)x(K+1) respectively. Let S' be the resulting matrix obtained by forcing the smallest singular value $\sigma_{K+1} = 0$ and let W be the matrix which is constructed in the same manner as Y with noise-free data. The matrix $Y' = US'V^*$ turns out to be the closest matrix of rank K to Y in the sense of Frobenius norm [11]. Because the rank of the matrix W is K, it is reasonable that the weakest eigenvector of Y^*Y is an estimate of coefficient vector x. But what statistical properties do the eigenvectors have? In particular can we say that the weakest eigenvector is an unbiased estimator of x? It was pointed out in [12] that

$$E[Y^*Y] = W^*W + I(N-K)\sigma^2 \tag{11}$$

where E is an expectation operator, I is the identity matrix, and σ^2 is a variance. The eigenvalue decomposition of (11) is

$$E[Y^*Y] = V[D + I(N-K)\sigma^2]V^* \tag{12}$$

where D is the diagonal matrix whose elements are the eigenvalues of W^*W and V is the modal matrix of W^*W. Note that the eigenvectors of W^*W are preserved and each eigenvalue is increased by $(N-K)\sigma^2$. This property can be used for order selection. But (12) itself cannot be a proof of the supposition that the expectation of the eigenvectors is unaffected by noise, i.e. unbiased, because the eigenvectors of $E[Y^*Y]$

and the average of eigenvectors of Y^*Y will be different. In fact a small perturbation on W^*W might give a large error of eigenvector estimation, especially when some eigenvalues are clustered together [13]. At present no proof is available to support the claim that the expectation of the eigenvectors is unaffected by noise. However, some simulation results show that pole estimates using SVD-based methods are much less biased than those using covariance method [14,15].

2.2 DEFLATION ALGORITHM

So far the matrix Y was taken to be (N-K)x(K+1). But the true (or appropriate) order of a system is usually unknown a priori. Constructing a data matrix as (N-M)x(M+1) and solving it to obtain an estimate results in M-K extraneous poles. It has been observed empirically that the presence of extraneous poles seems to protect the true poles against noise-induced perturbation [14,15]. Unfortunately it is not known how to select an optimum M given N and K without examining the error for each M.

If we overestimate the order, then the SVD of W (of order (N-M) x (M+1)) becomes

$$W = USV^* \tag{13}$$

where

$$S = \begin{bmatrix} \sigma_1 & & & 0 \\ & \sigma_2 & & \\ & & \ddots & \\ & & & \sigma_{M+1} \\ 0 & & & \\ & & 0 & \end{bmatrix} \begin{matrix} \\ \\ \\ \\ \} \\ \\ \end{matrix} N-2M-1$$

Here $\sigma_1 \geq \sigma_2 \geq \cdots \geq = \sigma_{K+1} = \sigma_{K+2} = \cdots = \sigma_{M+1} = 0$ and U and V are the unitary matrices of the appropriate orders. The last M-K+1 columns of V are in the null space of W. Let

$$H = [v_{K+1}, v_{K+2}, \cdots, v_{M+1}] \tag{14}$$

where v_i is the ith column of V.

The columns of H span the null space of W. Define an (M+1)x(M-K+1) matrix

$$G = \begin{bmatrix} b_K & 0 & & & \\ b_{K-1} & b_K & & & \\ \cdot & b_{K-1} & \cdot & & \\ \cdot & \cdot & \cdot & \cdot & \\ \cdot & \cdot & & \cdot & b_K \\ b_0 & \cdot & & & b_{K-1} \\ 0 & b_0 & & & \cdot \\ & & \cdot & & \cdot \\ & & & \cdot & \cdot \\ & \bigcirc & & & b_0 \end{bmatrix} \tag{15}$$

It is shown in appendix A that every column of G is in the null space of W so that columns of G are the linear combinations of columns of H and vice versa. By introducing (M-K+1)x(M-K+1) matrix R, G and H are related as follows.

$$G = HR \tag{16}$$

Given the SVD of Y, we can examine the σ_i's. If there is a significant drop between successive values of σ_m and σ_{m+1}, and if the singular values after σ_{m+1} are relatively small and there are no big changes between successive singular values, then the order can be selected as $\hat{K} = m$. Now any column vector of H is a candidate for a coefficient vector. The problem is how to estimate G efficiently, given H.

The essence of Henderson's deflation algorithm [16] is as follows.

1) Perform a forward Gauss elimination with partial pivoting to reduce H^T to an upper trapezoidal matrix, leaving a triangular pattern of zeros in the lower left corner.
2) Perform a backward Gauss elimination without pivoting to reduce a triangular pattern of zeros in the upper right while preserving those in the lower left.

This resulting matrix was suggested [16] as an approximate G^T except for row scaling. There is a problem with this algorithm. Because of the finite word length of a digital computer, the use of a backward Gauss elimination without pivoting may result a large error. This error may emphasize certain directions which are not dominant originally.

We suggest an algorithm that overcomes this difficulty. Let $g^{(m)}$ be the mth column of G and $r^{(m)}$ be the mth column of R, then from Equation (16)

$$Hr^{(m)} = g^{(m)} \qquad \text{for } m = 1, 2, \ldots, M-K \tag{17}$$

By partitioning, (17) is rewritten as

$$\begin{bmatrix} H_1 \\ \hline H_2 \\ \hline H_3 \end{bmatrix} r^{(m)} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ \hline b_K \\ \vdots \\ b_0 \\ \hline 0 \\ \vdots \\ 0 \end{bmatrix} \tag{18}$$

Because the (b_i) are unknown, $r^{(m)}$ can be estimated such that $r^{(m)}$ satisfies the following set of M-K homogeneous equations.

$$\begin{bmatrix} H_1 \\ \hline H_3 \end{bmatrix} r^{(m)} = 0 \tag{19}$$

Intentionally, (19) is changed to the set of inhomogeneous equations such that

$$\begin{bmatrix} H_{1L} \\ \hline H_{3L} \end{bmatrix} r_I^{(m)} = - \begin{bmatrix} h_{1R} \\ \hline h_{3R} \end{bmatrix} \tag{20}$$

where $r_I^{(m)}$ is the vector consisting of the first M-K elements of $r^{(m)}$, H_{1L} (H_{3L}) is the matrix consisting of the first M-K columns of H_1 (H_3) and h_{1R} (h_{3R}) is the last column of $H_1(H_3)$. Gauss elimination with (possibly) complete pivoting is used to solve Equation (20) for each m to get $\hat{R}$. Now

$$H\hat{R} = \hat{G} \tag{21}$$

$\hat{G}$ in (21) is intended to approximate G except for column scaling. From $\hat{G}$, a (K+1)x(M-K+1) matrix Q is obtained by eliminating the zeros and shifting every element of the kth column of $\hat{G}$ upward (k-1) times for k = 1, ..., M-K+1. Each column of Q should be an estimate of the coefficient vector. Taking an average of (M-K+1) columns of Q after normalizing each column with respect to the last element of that column is the final reduced estimate of the true coefficient vector. Gauss elimination with complete pivoting to estimate each $r^{(m)}$ should cure the problem of emphasizing a particular direction with Henderson's deflation algorithm. And averaging normalized columns of Q should alleviate the problem further. Simulation results for the examples chosen for Experiment I in Chapter 5 show that sample variances for coefficients with the new algorithm were always smaller than those with Henderson's deflation algorithm for K = 2 case. Also Experiment II in Chapter 5 shows that the bias with the new algorithm was relatively smaller than that with the existing method.

CHAPTER 3

ITERATIVE PREPROCESSING ALGORITHM

The problem of estimating poles can be posed in terms of a time-domain system identification problem. Suppose the unknown system has a K-th order transfer function given by

$$H(z) = \frac{A(z)}{B(z)} = \frac{a_0 + a_1 z^{-1} + \cdots + a_{K-1} z^{-(K-1)}}{b_0 + b_1 z^{-1} + \cdots + b_K z^{-K}} \tag{22}$$

$$= w_0 + w_1 z^{-1} + w_2 z^{-2} + \cdots\cdots$$

where $b_0 = 1$ for convenience and (w_n) are samples of the impulse response. This can be written in terms of the difference equation

$$\begin{cases} \sum_{i=0}^{n} b_i w_{n-i} = a_n, & n = 0, 1, \ldots, K-1 \\ \sum_{i=0}^{K} b_i w_{n-i} = 0, & n \geq K \end{cases} \tag{23}$$

or in matrix form (with N samples of the impulse response)

$$\begin{bmatrix} \cdot & \bigcirc & & & w_0 \\ & & & w_0 & w_1 \\ & & \cdot & & \cdot \\ & & \cdot & & \cdot \\ & \cdot & & & \cdot \\ w_0 & & w_1 & \cdots\cdots & w_K \\ w_1 & & w_2 & \cdots\cdots & w_{K+1} \\ \cdot & & \cdot & & \cdot \\ \cdot & & \cdot & & \cdot \\ w_{N-K-1} & & w_{N-K} & \cdots\cdots & w_{N-1} \end{bmatrix} \begin{bmatrix} b_K \\ \cdot \\ \cdot \\ b_1 \\ 1 \end{bmatrix} = \begin{bmatrix} a_0 \\ a_1 \\ \cdot \\ \cdot \\ a_{K-1} \\ 0 \\ 0 \\ \cdot \\ \cdot \\ 0 \end{bmatrix} \tag{24}$$

$$\tilde{W} \qquad x \qquad = \begin{bmatrix} a \\ 0 \end{bmatrix}$$

(24) can be rewritten as

$$\begin{bmatrix} 1 & & & & & \\ b_1 & 1 & & & & \\ \vdots & \vdots & \ddots & & & \\ b_K & \cdots & b_1 & 1 & & \\ & \ddots & & \ddots & \ddots & \\ & & b_K & \cdots & b_1 & 1 \end{bmatrix} \begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_{K-1} \\ w_K \\ \vdots \\ w_{N-1} \end{bmatrix} = \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_{K-1} \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{25}$$

$$B \qquad w \qquad = \begin{bmatrix} a \\ 0 \end{bmatrix}$$

We observe $y_n = w_n + e_n$ for n = 0, 1, ..., N-1, where (e_n) are samples of uncorrelated, zero-mean noise. In vector form

$$y = w + e \tag{26}$$

where $y = [y_0,\ldots,y_{N-1}]^T$ and $e = [e_0,\ldots,e_{N-1}]^T$. Given y we want to estimate the denominator coefficients in (22). Let $\tilde{Y}$ be the matrix which is constructed in the same manner as $\tilde{W}$ with (y_n) as elements. Then from (24) we have

$$\tilde{Y}x = \begin{bmatrix} a \\ 0 \end{bmatrix} + \tilde{d} \tag{27}$$

where $\tilde{d} = (\tilde{Y} - \tilde{W})x = Be$ is called the equation error vector. Now (27) can be rewritten as

$$Be = \tilde{Y}x - \begin{bmatrix} a \\ 0 \end{bmatrix} \tag{28}$$

Because Matrix B is lower triangular and has 1's as diagonal elements, it is nonsingular and invertible. The inverse of matrix B is easily obtained by forward substitution.

$$F = B^{-1} = \begin{bmatrix} 1 & & & & & \\ f_1 & 1 & & & & \\ f_2 & f_1 & 1 & & & \\ \cdot & \cdot & \cdot & \cdot & & \\ \cdot & \cdot & & \cdot & \cdot & \\ f_{N-2} & f_{N-3} & \cdots\cdots & f_1 & 1 & \\ f_{N-1} & f_{N-2} & \cdots\cdots & f_2 & f_1 & 1 \end{bmatrix} \tag{29}$$

where $f_0 = 1$

$$\text{and } f_j = \begin{cases} \sum_{i=1}^{j} b_1 f_{j-i}, & 1 \le j \le K \\ \sum_{i=1}^{K} b_1 f_{j=i}, & K+1 \le j . \end{cases}$$

Note that $1/B(z) = 1 + \sum_{i=1}^{\infty} f_i z^{-i}$.

By introducing the matrix F, (28) becomes

$$e = F[\tilde{Y}x - \begin{bmatrix} a \\ 0 \end{bmatrix}] = F\tilde{Y}x - F_L a \tag{30}$$

where F_L is the matrix consisting of first K columns of F. Let $\tilde{Y}_L$ be the matrix consisting of first K columns of $\tilde{Y}$ and $\tilde{y}_R$ be the $K+1^{th}$ column of $\tilde{Y}$. Note that $y = \tilde{y}_R$. Define $x_I = [b_K, b_{K-1}, \ldots, b_1]^T$. Then (9) becomes

$$e = F\tilde{Y}_L x_I + Fy - F_L a \tag{31}$$

Because F (and so F_L) is a function of (b_i), minimizing a sum of squared error $J = e^* e$ is unfortunately a highly nonlinear problem. But fixing F as a constant matrix and setting the derivatives of J with respect to x_I and a, we get two conditions:

$$F_L{}^* F_L\, a - F_L{}^* F\tilde{Y}_L\, x_I = F_L{}^* Fy \tag{32}$$

$$-(F\tilde{Y}_L)^* F_L\, a + (F\tilde{Y}_L)^* F\tilde{Y}_L\, x_I = -(F\tilde{Y}_L)^* Fy\ . \tag{33}$$

For a fixed F, the problem is linear in a and x_I. An iterative procedure can be used to minimize J. The first estimate of $a^{(1)}$ and $x_I{}^{(1)}$ results from taking $F^{(0)}$ as the identity matrix. With this $x_I^{(1)}$, we can construct $F^{(1)}$ to get a new estimate $a^{(2)}$ and $x_I{}^{(2)}$, and so forth. At each iteration, the previous denominator coefficients are used to get new estimates, so that $a^{(m)}$ and $x_I{}^{(m)}$ are found by solving a system of linear equations

$$\begin{bmatrix} (F_L^{(m-1)})^* F_L^{(m-1)} & -(F_L^{(m-1)})^* F^{(m-1)}\tilde{Y}_L \\ -(F^{(m-1)}\tilde{Y}_L)^* F_L^{(m-1)} & (F^{(m-1)}\tilde{Y}_L)^* F^{(m-1)}\tilde{Y}_L \end{bmatrix} \begin{bmatrix} a^{(m)} \\ x_I^{(m)} \end{bmatrix} = \begin{bmatrix} (F_L^{(m-1)})^* F^{(m-1)} y \\ -(F^{(m-1)}\tilde{Y}_L)^* F^{(m-1)} y \end{bmatrix}$$

for $m = 1, 2, \ldots,$ and $F^{(0)} = I$. (34)

If convergence is obtained, $J = e^* e$ is minimized, which is the error we want to minimize. But if the SNR is very low, it is hard to expect convergence in general. One can check that (34) is an explicit form of the Steiglitz-McBride iterative algorithm with an impulse input [5,6]. This method has been widely used and the convergence and accuracy properties for large data lengths and/or high SNR assumptions are well-known [17]. At each iteration, it is necessary to solve a system of 2K linear equations.

3.1 ITERATIVE PREPROCESSING ALGORITHM

Because the rank of the matrix F_L is K, the matrix $F_L^* F_L$ is hermitian positive definite and invertible. From (32), we get

$$a = (F_L^* F_L)^{-1} [F_L^* \tilde{FY}_L x_I + F_L^* Fy] \tag{35}$$

Substituting (35) into (33) we obtain

$$(\tilde{FY}_L)^* \tilde{Q} \tilde{FY}_L x_I = -(\tilde{FY}_L)^* \tilde{Q} Fy \tag{36}$$

where $\tilde{Q} = I - F_L (F_L^* F_L)^{-1} F_L^*$

With a fixed F, this equation is linear in x_I. An iterative procedure analogous to that mentioned above is used to solve (36). The first estimate of $x_I^{(1)}$ results from setting $F^{(0)} = I$. With this $x_I^{(1)}$ we can construct $F^{(1)}$ and calculate $[F_L^{(1)*} F_L^{(1)}]^{-1}$ to get new estimate $x_I^{(2)}$, and so forth.

$$(F^{(m-1)}\tilde{Y}_L)^* \tilde{Q}^{(m-1)}(F^{(m-1)}\tilde{Y}_L)\, x_I^{(m)} = -(F^{(m-1)}\tilde{Y}_L)^* \tilde{Q}^{(m-1)}F^{(m-1)}y \tag{37}$$

where $\tilde{Q}^{(m-1)} = I - F_L^{(m-1)}(F_L^{(m-1)*}F_L^{(m-1)})^{-1}(F_L^{(m-1)})^*$

for $m = 1, 2, \ldots,$ and $F^{(0)} = I$.

At each iteration we need to solve a system of only K linear equations which is an advantage compared to the S-M algorithm. If convergence is obtained, then again the mean-square error is minimized.

Before proceeding further let us introduce the Evans-Fischl algorithm [18]. Partitioning the matrices in (28), we get

$$De = Yx = d \tag{38}$$

where D is the matrix of the last N-K rows of B, and Y and d are as in (7). The error we would like to minimize is e^*e. Because the matrix D is (N-K)xN, we cannot solve for e directly. But using the SVD

$D = USV^*$ ($DD^* = U(SS^*)U^*$) .

Thus, $e = VS^+U^*d$ which is the minimum norm solution so that

$$\min e^*e = \min d^*U(SS^*)^{-1}U^*d = \min d^*(DD^*)^{-1}d$$

i.e.,

$$\min e^*e = \min (Y_L x_I + y_R)^*(DD^*)^{-1}(Y_L x_I + y_R) \tag{39}$$

where Y_L and y_R are defined as in (8). With a fixed DD^*, the minimum of e^*e would be given by the solution for x_I from the normal equation

$$Y_L (DD^*)^{-1} Y_L x_I = -Y_L^* (DD^*)^{-1} y_R \qquad (40)$$

Since D does in fact depend on x_I, we can use (40) but with iterations. It is shown in Appendix B that the iterative preprocessing algorithm, i.e., Equation (37) is basically the same as the Evans-Fischl algorithm. At each iteration an (N-K)x(N-K) matrix inversion is required in the E-F algorithm, but inversion of KxK matrix is required for the IPA. Readers might notice that the matrix multiplication $\tilde{FY}$ is nothing but an inverse filtering and it requires only about $NK-K^2/2$ multiplications. When N is very large, the IPA which avoids inversion of a large matrix ((N-K)x(N-K)) should be more stable than the E-F algorithm. For large N, about $(N^3 + N^2K)/6$ multiplications except for solving a system of K linear equations are required at each iteration of the E-F algorithm. The Equation (36) is rewritten as

$$[(\tilde{FY})^*(\tilde{FY}) - (F_L^* \tilde{FY})^* (F_L^* F_L)^{-1} (F_L^* \tilde{FY})]\, x = 0.$$

As mentioned before $NK-K^2/2$ multiplications are required to compute each of F and $\tilde{FY}$. To obtain $(\tilde{FY})^*(\tilde{FY})$, $NK^2/2$ multiplications are required. Because of the special forms of F_L and $\tilde{FY}$, only 2NK is required to compute $(F_L^* \tilde{FY})$. Thus, including an inversion of $(F_L^* F_L)$, about $(3K^3/2 + 4NK + NK^2/2)$ multiplications except solving a system of K linear equations are required at each iteration of the IPA which is a great advantage compared to the E-FA especially when N >> K.

3.2 APPROXIMATE ITERATIVE PREPROCESSING ALGORITHM

The disadvantage of the IPA is that a KxK matrix inversion is required at each iteration. In at least two cases of interest we can avoid such a matrix inversion.

3.2.1 Damped Sinusoid Case

If the signal consists of damped sinusoids, i.e. each $|z_i| < 1$ for $i = 1, 2, \ldots, K$, then from a practical viewpoint the signal has finite duration. This observation can be used to simplify the calculation of $(F_L^* F_L)^{-1}$. As a result both the computational load and the stability of the IPA are improved. Suppose the B matrix is partitioned as

$$B = \begin{array}{c} \\ K \\ K \\ N-2K \end{array} \left[\begin{array}{c|c} B_0 & 0 \\ \hline B_1 & \\ \hline & B_2 \\ 0 & \end{array}\right] \begin{array}{l} \\ \\ \\ \} N-K \end{array} \qquad (41)$$

$$\begin{array}{cc} K & N-K \end{array}$$

Let $R = \lim_{N\to\infty} (F_L^* F_L)$. It is shown in Appendix C that R^{-1} relates to the matrix B very simply as

$$R^{-1} = B_0 B_0^* - B_1^* B_1 \qquad (42)$$

Approximating $(F_L^* F_L)^{-1}$ for finite N by R^{-1}, (36) becomes

$$(\tilde{FY_L})^* \tilde{Q}\, \tilde{FY_L}\, x_I = -(\tilde{FY_L})^* \tilde{Q}\, Fy \qquad (43)$$

where $\tilde{Q} = I - F_L R^{-1} F_L^*$

Convergence and accuracy properties are not available yet, but simulation results for the examples chosen in Chapter 5 show that this approximate iterative preprocessing algorithm (AIPA) converges to the maximum-likelihood estimator for moderate SNR.

3.2.2 Pure Sinusoid Case

It is shown in Appendix D that $\lim_{N\to\infty} (F_L^* F_L)^{-1} = 0$ when the signal consists of pure sinusoids, that is when the roots of $B(z) = 0$ are on the unit circle. Because

$$\lim_{N\to\infty} a = \lim_{N\to\infty} (F_L^* F_L)^{-1}[F_L^* \tilde{FY}_L x_I - F_L^* Fy] = 0 ,$$

(33) becomes

$$(\tilde{FY}_L)^* \tilde{FY}_L x_I = -(\tilde{FY}_L)^* Fy \qquad (44)$$

These equations are the same as Kay's IFA (iterative filtering algorithm) [7]. Our derivation emphasize that (44) is an approximation which is good when $N \to \infty$. We show in Appendix B that (36) can be rewritten as .

$$(F_2 Y_L)^* Q F_2 Y_L x_I = -(F_2 Y_L)^* Q F_2 y_R \qquad (45)$$

where $Q = I - F_1(F_L^* F_L)^{-1} F_1^*$

Here Y_L and Y_R are defined as in (8) and

$$F = \left[\begin{array}{c|c} F_0 & 0 \\ \hline F_1 & F_2 \end{array}\right] \begin{matrix} \} K \\ \} N-K \end{matrix} \qquad F_L = \left[\begin{array}{c} F_0 \\ \hline F_1 \end{array}\right] .$$

For large N we can use $(F_L^* F_L)^{-1} = 0$ to approximate (45) as

$$(F_2 Y_L)^* (F_2 Y_L)\, x_I = -(F_2 Y_L)^* F_2 y_R \tag{46}$$

Equation (46) is the same as (44) when $N = \infty$. Note that the matrix multiplication $F_2 Y$ is different from the matrix multiplication $F\tilde{Y}$, because the former filters the columns of Y separately and the latter filters the data $(y_0, y_1, \ldots, y_{N-1})$.

Equation (30) for the error is rewritten as

$$e = F_L (Y_0 x - a) + F_2 Y x \tag{47}$$

where Y_0 is the matrix of the first K rows of $\tilde{Y}$. To minimize $e^* e$ given $a = Y_0 x$ with a fixed F_2, we need to solve the set of normal equations which is identical to (46) with iterations. Both Equations (44) and (46) are good when $N = \infty$. Equation (44) comes from setting $a = 0$. This corresponds to setting the initial conditions in the difference equation to zero. If we use Equation (44) to estimate frequencies when N is small, $a = 0$ is not appropriate. This explains why Kay [7] got a large sample variance for the frequency estimates with large SNR. On the other hand Equation (46) assumes $a = Y_0 x$, i.e. sets initial conditions to $y_0, y_1, \ldots, y_{K-1}$, which is more reasonable especially when N is small. Because (46) filters the columns of Y separately, it requires an additional $(N-K)\cdot(K-1)^2$ multiplications compared to Kay's method.

CHAPTER 4

CRAMER-RAO LOWER BOUND

The previous chapter described an algorithm for estimating natural frequencies. We do not know how to describe generally the convergence property of the IPA. Simulation results using the IPA can be compared with simulation results using other methods. But is is more interesting to compare with the Cramer-Rao lower bound.

In this chapter we shall describe a calculation of the C-R bound for the coefficients of the characteristic equation. The maximum-likelihood estimate is an unbiased estimate whose conditional variance satisfies the C-R bound with equality. As mentioned in the introduction if the IPA converges to the ML estimate for the characteristic equation coefficients then it is maximum-likelihood for the natural frequencies as well. The C-R bound is stated as follows [1]. Let y be an observation vector and $\hat{\theta}$ be any unbiased estimate of a vector θ, then the conditional variance is bounded by

$$\text{var}(\hat{\theta}_i | \theta_i) \geq [J^{-1}]_{ii} \tag{48}$$

where the matrix J is

$$J = - E\left[\frac{\partial}{\partial \theta} \left\{ \frac{\partial}{\partial \theta} \log p(y|\theta) \right\}^T\right] \tag{49}$$

and $p(y\ \theta)$ is the probability density function for the observation given the unknown parameter vector.

4.1 C-R BOUND FOR COEFFICIENT ESTIMATION

In coefficient estimation case, the model is either

$$w = -\tilde{W}_L x_I + \begin{bmatrix} a \\ 0 \end{bmatrix} . \tag{50}$$

or from (25)

$$w = F_L a \tag{51}$$

The observation is

$$y = w + e \tag{52}$$

where now we assume that e is a white Gaussian noise vector with variance σ^2.

Let $\theta = [b_K, b_{K-1}, \ldots, b_1, a_0, \ldots, a_{K-1}]^T$, then the conditional density function of y is

$$p(y|\theta) = (2\pi\sigma^2)^{-N/2} \exp[-\frac{1}{2\sigma^2}(y-w)^T(y-w)] \tag{53}$$

To evaluate the bound we must compute the second partial derivative of $\log p(y|\theta)$ with respect to θ. First we will evaluate $\frac{\partial w}{\partial \theta}$. Because F_L in the Equation (51) is not a function of a, $\frac{\partial w}{\partial a}$ is simply

$$[\frac{\partial w}{\partial a}] = F_L \tag{54}$$

Rewrite Equation (50) as

$$\begin{bmatrix} w_0 \\ w_1 \\ \cdot \\ \cdot \\ w_{K-1} \\ w_K \\ w_{K+1} \\ \vdots \\ w_{N-1} \end{bmatrix} = - \begin{bmatrix} 0 & 0 & \cdots & 0 \\ & \bigcirc & & w_0 \\ & & \cdot & \cdot \\ & & \cdot & \cdot \\ 0 & w_0 & \cdots & w_{K-2} \\ w_0 & w_1 & \cdots & w_{K-1} \\ w_1 & w_2 & \cdots & w_K \\ \vdots & \vdots & & \vdots \\ w_{N-K-1} & w_{N-K} & \cdots & w_{N-2} \end{bmatrix} \begin{bmatrix} b_K \\ \cdot \\ \cdot \\ \cdot \\ b_1 \end{bmatrix} + \begin{bmatrix} a_0 \\ a_1 \\ \cdot \\ \cdot \\ a_{K-1} \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{55}$$

$$w = - \tilde{W}_L x_I + \begin{bmatrix} a \\ 0 \end{bmatrix}$$

We claim that

$$\left[\frac{\partial w}{\partial x_I}\right] = - F\tilde{W}_L \tag{56}$$

Let $I^{(m)}$ = mth row of the identity matrix,

$F^{(m)}$ = mth row of the matrix F, and

$\tilde{W}_L{}^{(m)}$ = mth column of the matrix $\tilde{W}_L$.

Then from (29) it is easy to show that

$$- \sum_{i=1}^{K} b_i F^{(m-i)} = F^{(m)} - I^{(m)} \tag{57}$$

And it is obvious that

$$I(m)\tilde{}_W(j) = \begin{cases} w_{m+j-(K+2)} & \text{for } m+j \geq K+2 \\ 0 & \text{for } m+j < K+2 \end{cases} \qquad (58)$$

We will use induction to show that (56) holds. From (55), it is clear that

$$\frac{\partial w_i}{\partial b_i} = -w_0 \quad \text{and} \quad \left.\frac{\partial w_m}{\partial b_i}\right|_{m<i} = 0 \qquad (59)$$

for $i = 1, 2, \ldots, K$.

Thus, Equation (59) satisfies Equation (56). Suppose Equation (56) holds for $m = 0, 1, \ldots, M$, i.e.

$$\frac{\partial w_m}{\partial b_i} = -F(m+1)\tilde{}_{W_L}(K-i+1) \qquad (60)$$

for $i = 1, 2, \ldots, K$ and $m = 0, 1, \ldots, M$.

Then for $m = M + 1$,

$$\frac{\partial w_{M+1}}{\partial b_i} = \frac{\partial}{\partial b_i}\left[\sum_{j=0}^{K} -b_j w_{M+1-j}\right] \qquad (61)$$

$$= -w_{M+1-i} - \sum_{j=1}^{K} b_j \frac{\partial w_{M+1-j}}{\partial b_i}$$

$$= -w_{M+1-i} + \sum_{j=1}^{K} b_j F(M+2-j)\tilde{}_{W_L}(K-i+1) \qquad \text{by (60)}$$

$$= - w_{M+1-i} - F^{(M+2)}\tilde{W}_L\,(K-i+1) + I^{(M+2)}\tilde{W}_L\,(K-i+1) \quad \text{by} \quad (57)$$

$$= - F^{(M+2)}\tilde{W}_L\,(K-i+1) \quad \text{by} \quad (58)$$

for $i = 1, 2, \ldots, K$.

By induction (56) holds for any $i = 1, 2, \ldots, K$ and $m = 0, 1, \ldots, N-1$.

Thus, from (54) and (56)

$$\left[\frac{\partial w}{\partial \theta}\right] = [-F\tilde{W}_L \,\vdots\, F_L] \tag{62}$$

so that

$$\frac{\partial}{\partial \theta} \log p(y|\theta) = \frac{1}{\sigma^2}\,[-F\tilde{W}_L \,\vdots\, F_L]^T(y-w) \tag{63}$$

Let $G(\theta) = [-F\tilde{W}_L \,\vdots\, F_L]$, then

$$\frac{\partial}{\partial \theta}\left[\frac{\partial}{\partial \theta} \log p(y|\theta)\right]^T \tag{64}$$

$$= \frac{1}{\sigma^2}\left[\frac{\partial}{\partial \theta}\,[(y-w)^T G(\theta)]\right]$$

$$= - \frac{1}{\sigma^2}\left[\left[\frac{\partial w}{\partial \theta}\right]^T G(\theta) - (y-w)^T \frac{\partial}{\partial \theta} G(\theta)\right]$$

because $E[(y-w)] = 0$ and $\left[\frac{\partial w}{\partial \theta}\right] = G(\theta)$,

$$J = -E\left\{\frac{\partial}{\partial \theta}\left[\frac{\partial}{\partial \theta} \log p(y|\theta)\right]^T\right\} \tag{65}$$

$$= \frac{1}{\sigma^2} G^T(\theta)G(\theta)$$

$$= \frac{1}{\sigma^2}\begin{bmatrix} (\tilde{FW}_L)^T\tilde{FW}_L & -(\tilde{FW}_L)^T F_L \\ -F_L{}^T\tilde{FW}_L & F_L{}^T F_L \end{bmatrix}$$

Finally,

$$J^{-1} = \sigma^2 \begin{bmatrix} P & Q \\ Q^T & R \end{bmatrix} \tag{66}$$

What we are interested in is the C-R bound for the denominator coefficients $\theta_1, \theta_2, \ldots, \theta_K$. It is easy to show that

$$P = [(\tilde{FW}_L)^T[I-F_L(F_L{}^T F_L)^{-1}F_L{}^T]\tilde{FW}_L]^{-1} \tag{67}$$

so that

$$\text{var}(\hat{\theta}_i - \theta_i) \geq \sigma^2 P_{ii}$$

where $\theta_i = b_{K-i+1}$ for $i = 1, 2, \ldots, K$.

Note that the C-R bound is related to the noise free data and the coefficients of the characteristic equation. The matrix P^{-1} is the matrix of the normal equation which is solved at each iteration of the IPA.

4.2 C-R BOUND FOR POLE ESTIMATION

This section gives the C-R lower bound for the damping constant and frequency of a transient signal composed of one conjugate pole pair. Let observation

$$y_n = c_1 e^{j\beta_1} \exp[(-\alpha_1 + j2\pi f_1)n] + c_1 e^{-j\beta_1} \exp[(-\alpha_1 - j2\pi f_1)n] + e_n$$

for $n = 0, 1, \ldots, N-1$

where $c_1 e^{\pm j\beta_1} \equiv$ residue, $\alpha_1 \equiv$ damping constant, and $f_1 \equiv$ frequency. Also let $y = [y_0, y_1, \ldots, y_{N-1}]^T$ and $\theta = [f_1, c_1, \beta_1, \alpha_1]^T$.

Then, the probability density function for y given θ is written as

$$p(y|\theta) = (2\pi\sigma^2)^{-N/2} \exp\left[-\frac{1}{2\sigma^2} \sum_{n=0}^{N-1} \left| y_n - 2c_1 e^{-\alpha_1 n} \cos(\beta_1 + 2\pi f_1 n) \right|^2\right]$$

It is easy to show that $J = -E\left[\frac{\partial}{\partial\theta}\left\{\frac{\partial}{\partial\theta} \log p(y|\theta)\right\}^T\right]$

is expressed as

$$J = \frac{2}{\sigma^2} \begin{bmatrix} 4\pi^2 c_1^2 p_{2,-c} & -2\pi c_1 p_{1,s} & 2\pi c_1^2 p_{1,-c} & 2\pi c_1^2 p_{2,s} \\ & p_{0,+c} & -c_2 p_{0,s} & -c_1 p_{1,+c} \\ \text{Symmetric} & & c_1^2 p_{0,-c} & c_1^2 p_{1,s} \\ & & & c_1^2 p_{2,+c} \end{bmatrix}$$

where

$$p_{i,-c} = \sum_{n=0}^{N-1} n^i e^{-2\alpha n}[1-\cos 2(\beta_1+2\pi f_1 n)]$$

$$p_{i,+c} = \sum_{n=0}^{N-1} n^i e^{-2\alpha n}[1+\cos 2(\beta_1+2\pi f_1 n)]$$

$$p_{i,s} = \sum_{n=0}^{N-1} n^i e^{-2\alpha n} \sin 2(\beta_1+2\pi f_1 n)$$

for i = 0, 1, 2.

The C-R lower bound is given by

$$\text{var}\,(\hat{\theta}_i-\theta_i) \geq [J^{-1}]_{ii}$$

for i = 1, 2, 3, 4.

Define the % error by

$$\% \text{ error} = \frac{\sqrt{\text{var}(\hat{\theta}_i-\theta_i)}}{\hat{\theta}_i} \cdot 100$$

It is interesting to see how the % error is affected by different damping constants and frequencies for a given noise standard deviation. Using the C-R lower bound, we can bound the % error for various α_1 and f_1 values.

Consider the signal

$$y_n = \exp(-\alpha_1 nT)\cos(2\pi f_1 nT) + e_n$$

for n = 0, 1, ..., N-1.

A large enough number of samples N is chosen so that $y_n \approx 0$ for $n \geq N$.

The % error in α_1 and in f_1 is plotted for the following 3 cases.

1) f_1 is varied from 0.01 to 0.49 while $\alpha_1 = 0.1$ and T = 1. The % error plot is in Figure 1. The minimum error for both α_1 and f_1 is obtained when $f_1 \approx 0.45$. Notice that the errors decreases as f_1 increases until f_1 gets near the Nyquist rate, $f_1 = 0.5$ Hz.

2) α_1 is varied from 0.005 to 0.25 while $f_1 = 0.4$ and T = 1. From Figure 2(a), the % error is small when the damping constant α_1 is small. The smaller the damping constant the longer the duration of the signal. Thus, the estimation error should be smaller too. But when the number of samples is fixed as in Figure 2(b) the error is relatively large when $\alpha_1 < 0.03$.

3) In this case, T is varied from 0.01 to 0.95 while $\alpha_1 = 0.25$ and $f_1 = 0.5$. Figure 3(a) shows that when T is small the error is small. But when N is given as 30, 40, and 50, the behavior is similar to case 2. In each case as T approaches $2f_1 = 1$ = the Nyquist rate, the error increases rapidly. Figure 3(a) shows that if the whole waveform is used, i.e. the record length is several time constants long, then we should make the sampling interval as small as possible.

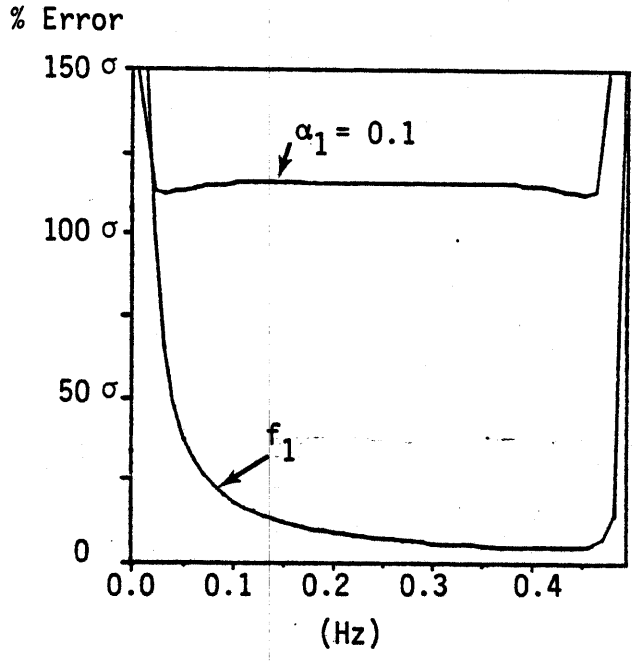
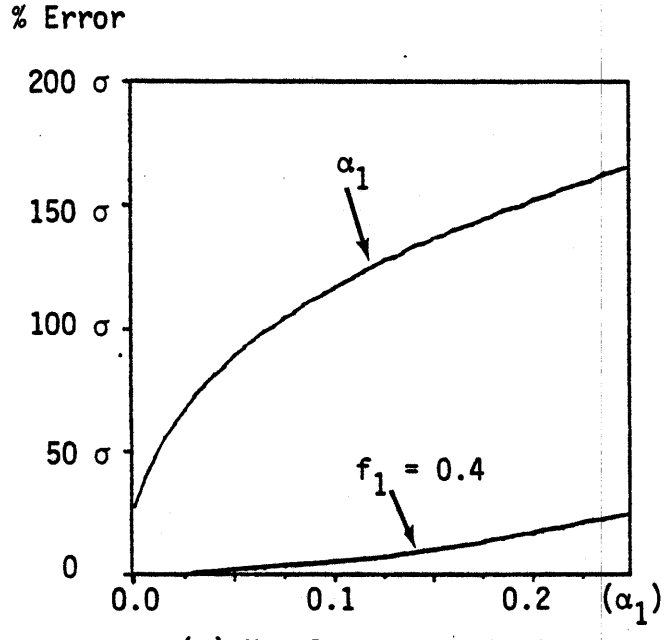
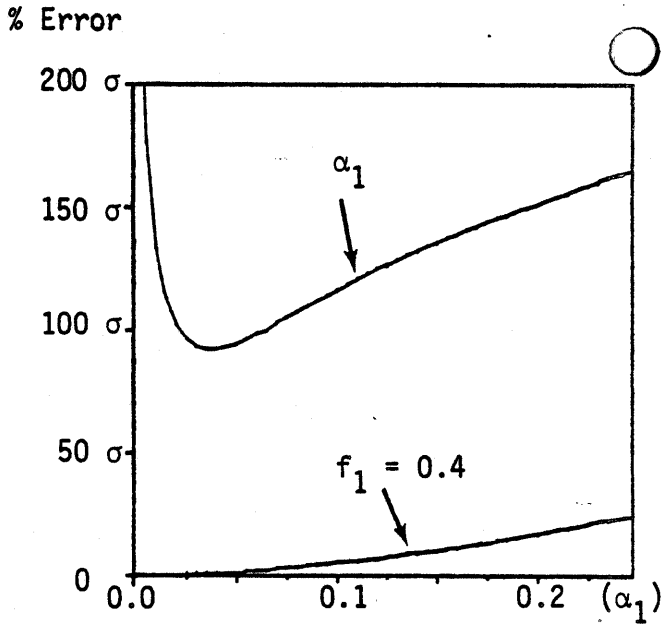


Figure 1. % error of α_1 and f_1 (α_1 is always 0.1).

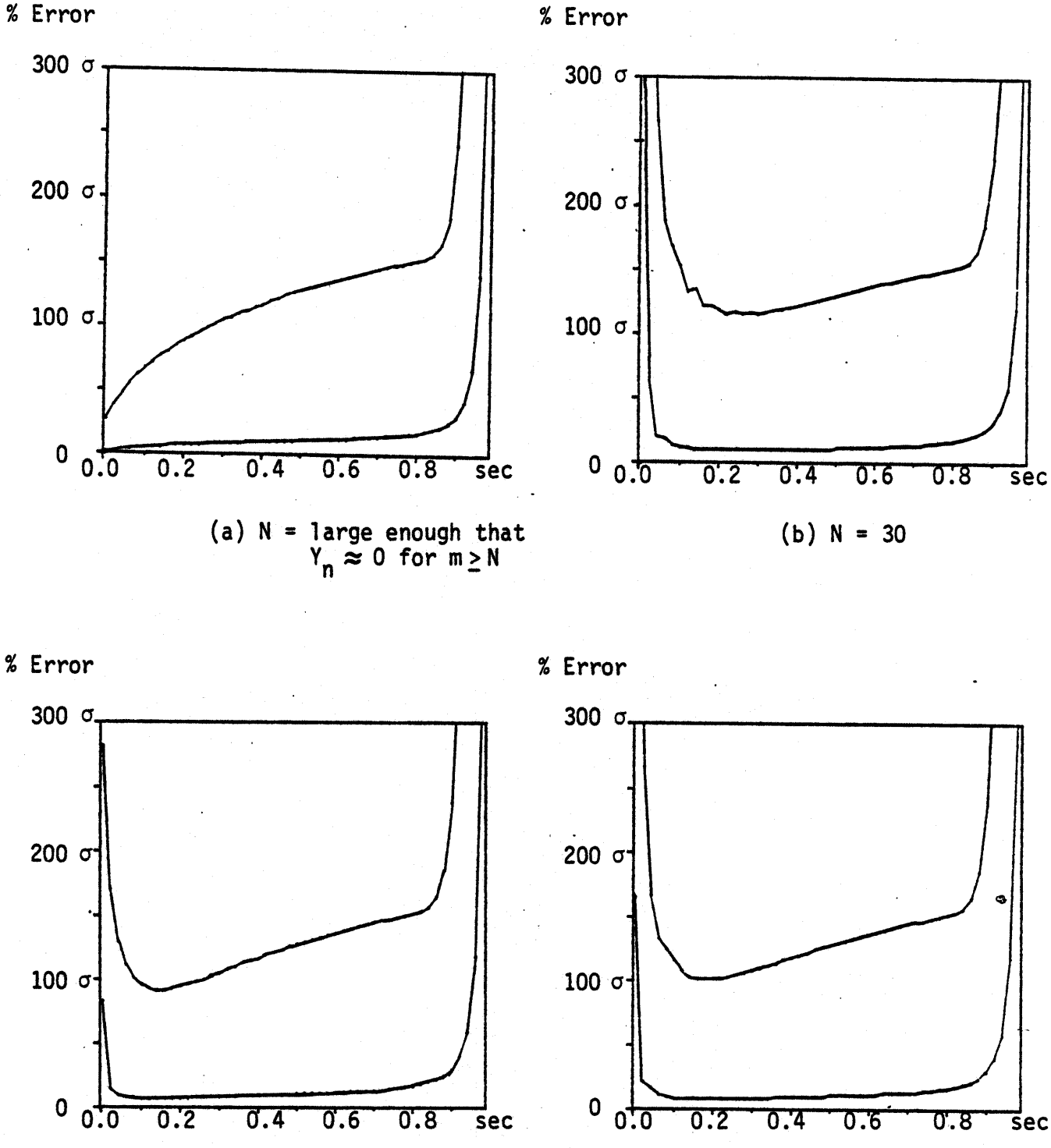


(a) N = large enough that $Y_n \approx 0$ for $n \geq N$



(b) N = 50

Figure 2. % error of α_1 and f_1 (f_1 is always 0.4).



(a) N = large enough that $Y_n \approx 0$ for $m \geq N$

(b) N = 30

(c) N = 40

(d) N = 50

Figure 3. % error of α_1 and f_1 (for different sampling period).

CHAPTER 5

SIMULATION RESULTS

In this chapter we give two kinds of experimental results to support the algorithms proposed in previous chapters.

5.1 EXPERIMENT I

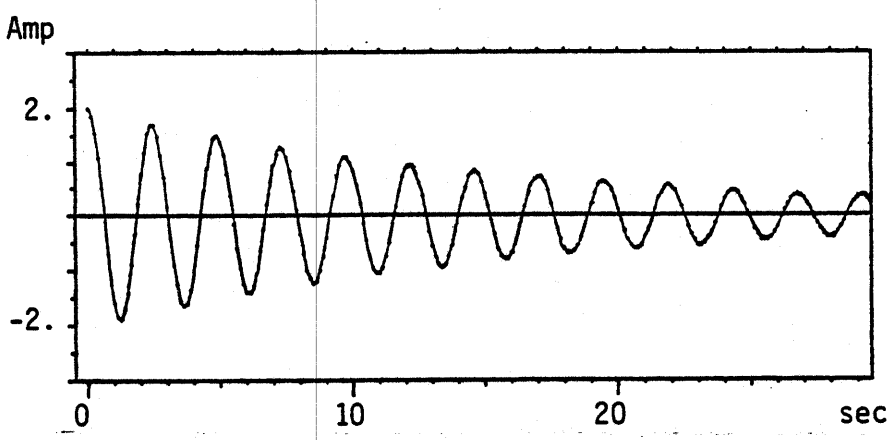
First we consider two data sets consisting of a pair of complex conjugate damped sinusoids in white Gaussian noise.

$$y_n = c_1 z_1^n + \bar{c}_1 \bar{z}_1^n + e_n \qquad \text{for } n = 0, 1, \ldots, N-1$$

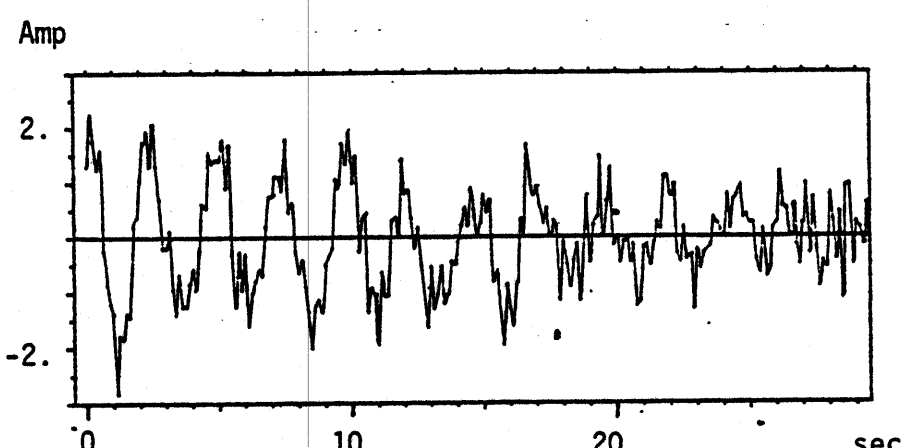
The first data set has $c_1 = 1$ and $z_1 = -0.8 + j0.5$ so that

$$H_1(z) = \frac{2 + 1.6z^{-1}}{1 + 1.6z^{-1} + 0.89z^{-2}}$$

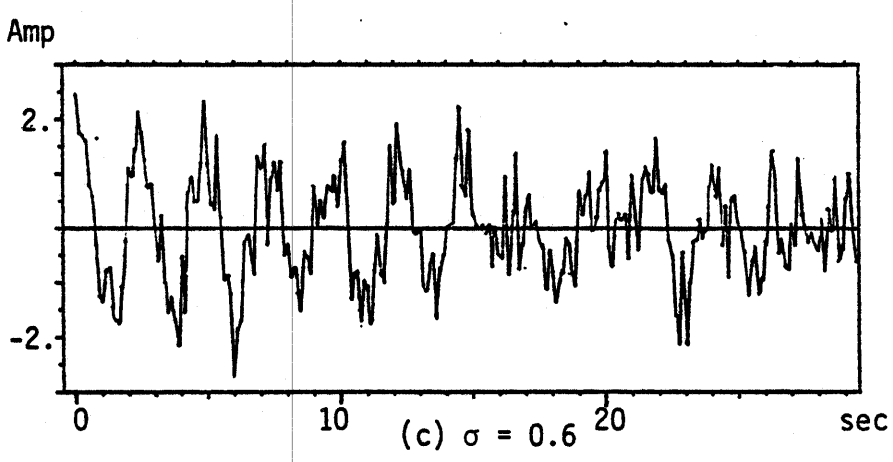
Waveforms with and without noise are plotted in Figure 4. The spread in the waveform due to noise is about 2σ where σ is the standard deviation of the additive noise e_n. Sampling rate T was 1 sec. SNR is defined as SNR = $[\text{peak}(y_n)/\sigma]^2 = 4/\sigma^2$. C-R bounds for $b_1 = 1.6$ and $b_2 = 0.89$ were calculated by the method suggested in Chapter 4. The variances with respect to the true coefficients were calculated after 500 trials at different SNR's for each method. The data record length N was 30 so that the AIPA is very close to the IPA. A 21x10 data matrix was used for the SVD method. As shown in Figure 1, both the AIPA and the S-M algorithm performed as maximum likelihood estimators up to 16 of SNR. But when the SNR was 11, the AIPA definitely outperformed the S-M algorithm. This indicates that the AIPA was more stable than the S-M



(a) $\sigma = 0.0$



(b) $\sigma = 0.5$



(c) $\sigma = 0.6$

Figure 4. Damped sinusoids plus white noise (I).

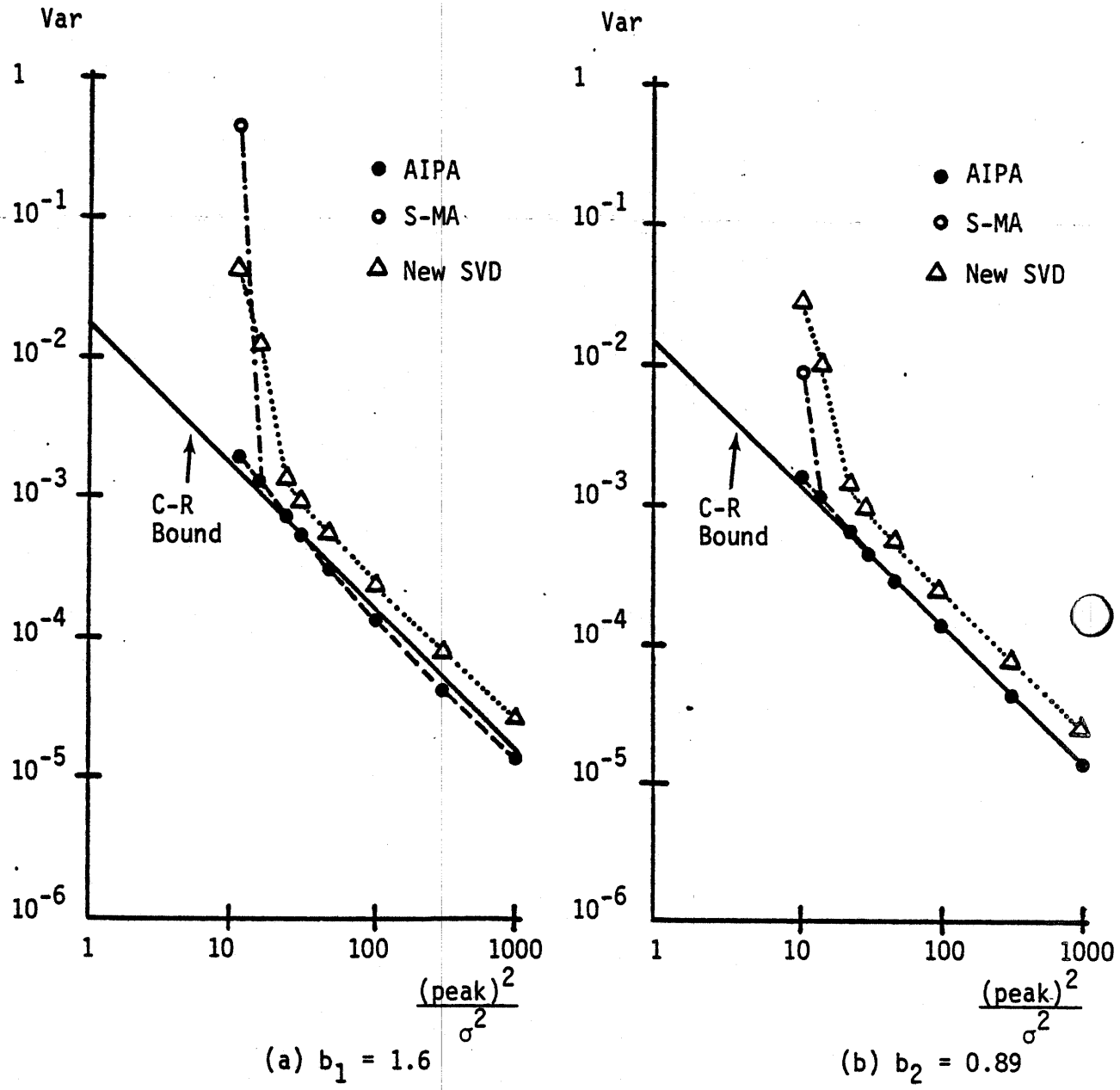


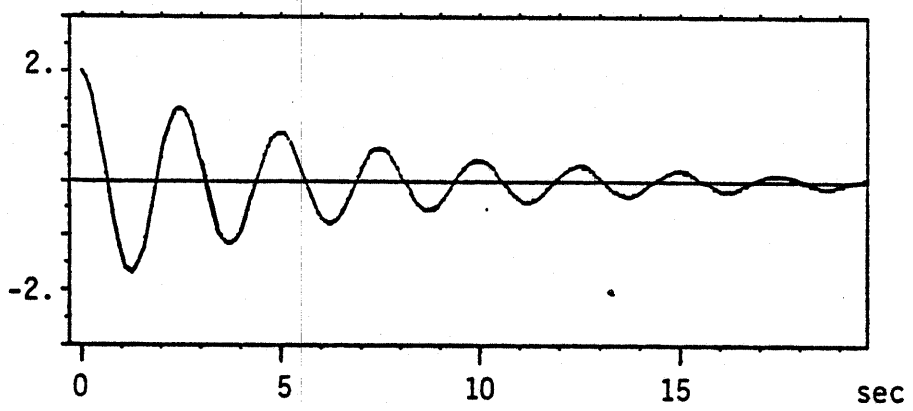
Figure 5. (a),(b) Comparison of pole estimators. ($b_1 = 1.6$, $b_2 = 0.89$)

algorithm. For the SVD method, eigenvectors corresponding to the 8 weakest eigenvalues were used to estimate the coefficient vector as described in Chapter 2. The sample variance of coefficients obtained by the SVD method was about the double the C-R bound.

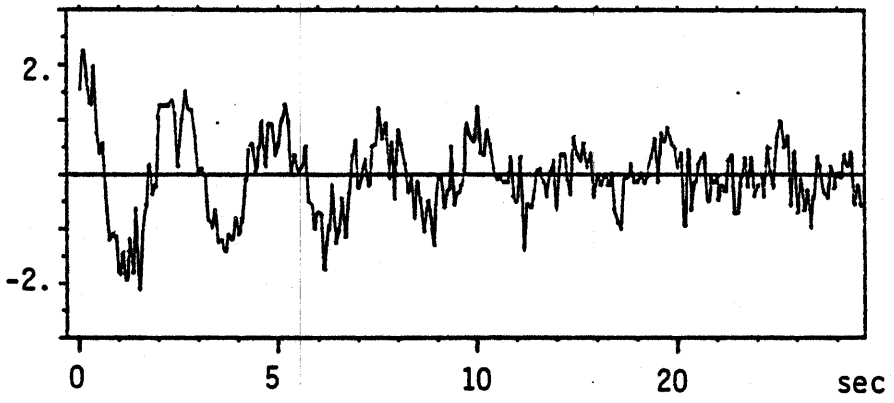
As a second example, $c_1 = 1$ and $z_1 = -0.7 + j0.5$ were used. Waveforms with and without noise are plotted in Figure 6. Sampling period T was 1 sec. The SNR is defined as before and N = 20. Again 500 trials were performed at different SNR for each method.

$$H_2(z) = \frac{2 + 1.4z^{-1}}{1 + 1.4z^{-1} + 0.74z^{-2}}$$

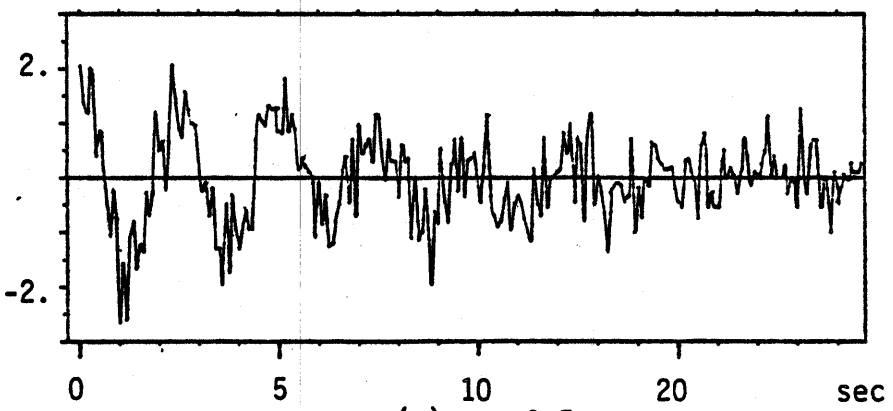
As shown in Figure 7, results from both the AIPA and S-M algorithm are poor below SNR = 44. But still we can see that the AIPA performs better than S-M algorithm for low SNR. A 14x7 data matrix was used for the SVD method. Again the sample variances for the coefficients obtained by the SVD method was about the double the C-R bounds. It should be noted that the new SVD method that combines some weakest eigenvectors always performed better than Henderson's deflation algorithm. Because the difference was about 1-2 dB for both examples, we do not show that on the graph. For higher order cases, we would expect to see a bigger difference. As mentioned in the introduction, if the estimator is the maximum-likelihood estimator for the coefficient vector then pole estimates via this estimator are also maximum-likelihood. This notion is demonstrated in the next examples.



(a) $\sigma = 0.0$



(b) $\sigma = 0.4$



(c) $\sigma = 0.5$

Figure 6. Damped sinusoids plus white noise (II).

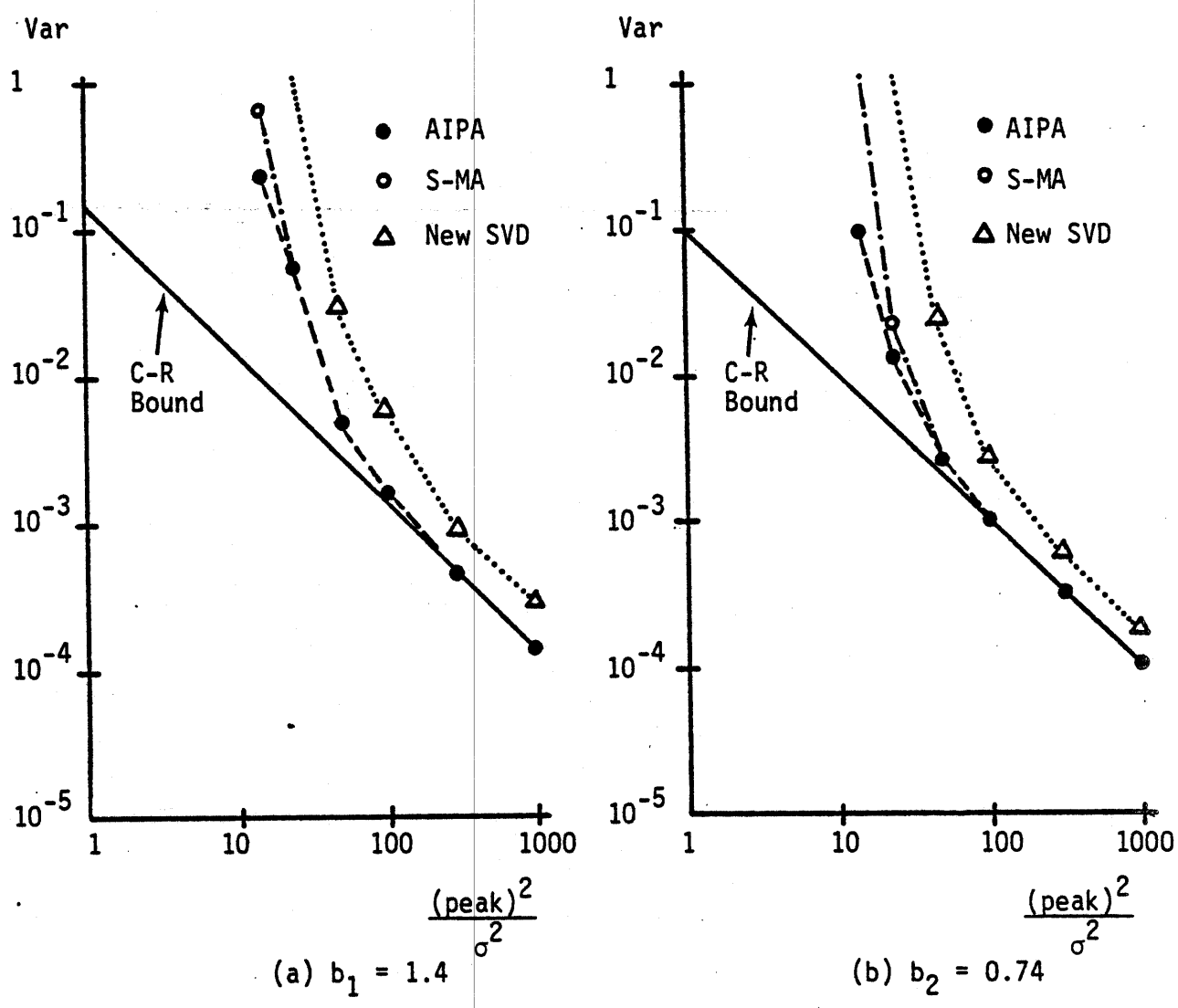


Figure 7. (a),(b) Comparison of pole estimators. (b_1 = 1.4, b_2 = 0.74)

5.2 EXPERIMENT II

In this experiment we used three data sets consisting of complex damped sinusoids in white Gaussian noise.

$$y_n = \sum_{i=1}^{K} \exp(s_i n) + e_n \qquad n = 0, 1, \ldots, 24$$

where $s_i = -\alpha_i + j2\pi f_i$

These three data sets were chosen by Kumaresan and Tufts [14]. The sequence e_n is white, and complex Gaussian, with variance $2\sigma^2$. SNR is $1/2\sigma^2$. The IPA was used to estimate the coefficient vector and then the pole damping factors (α_i) and pole frequencies (f_i) were calculated for each trial. The SVD method described in Chapter 2 was used for the K = 1 case. The C-R bounds as well as the backward covariance method using the SVD results have been extracted from [14]. For K = 1, $f_1 = 0.52$ was used. As shown in Figure 8, for $\alpha_1 = 0.1$, the IPA indeed performed as a maximum-likelihood estimator up to 10 of SNR. The new SVD method performed as well as K-T's method. An 18x8 data matrix was used for the new SVD method compared to 7x18 data matrix for K-T's method which means computing time for the new SVD method is much less. Note that the sample variance of the estimates in this one pole case show broad minimum somewhere between M = 6 to 10 for the new SVD method. For the results in Figure 9, $\alpha_1 = 0.2$ was used. Again the IPA performed as a maximum-likelihood estimator up to 30 of SNR. The sample variance using the new SVD method was larger than that for K-T's method. But as Table 1 shows, the bias for α_1 is smaller with the new SVD method. Averaging reduced coefficient vectors as mentioned in Chapter 2 might have reduced the bias.

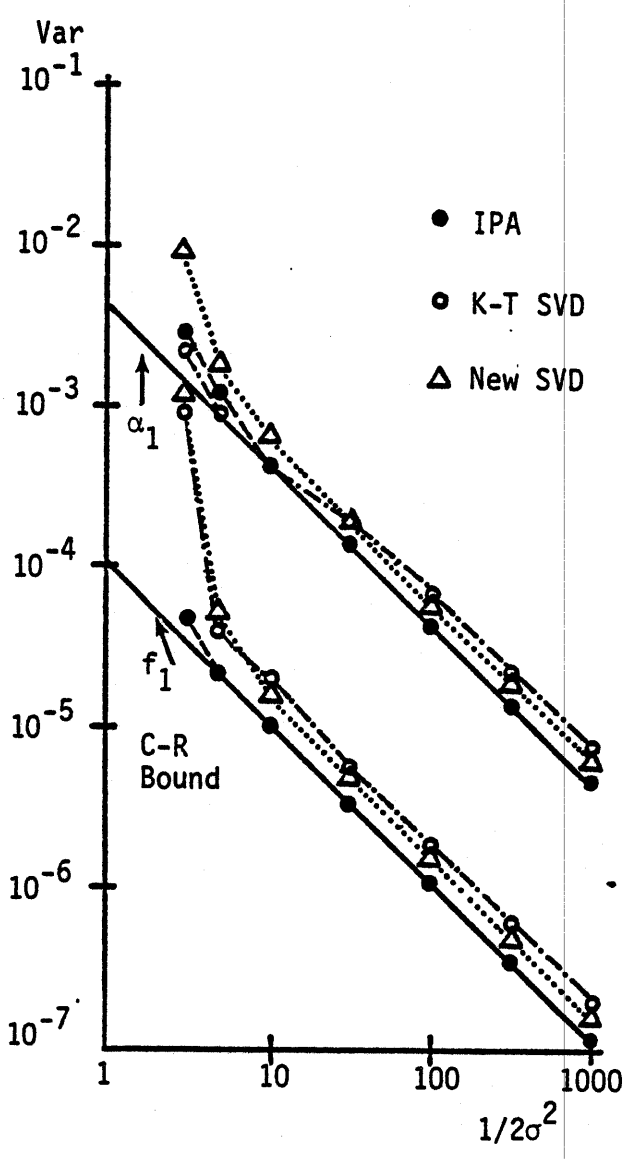


Figure 8. Comparison of pole estimators.

($\alpha_1 = 0.1$, $f_1 = 0.52$)

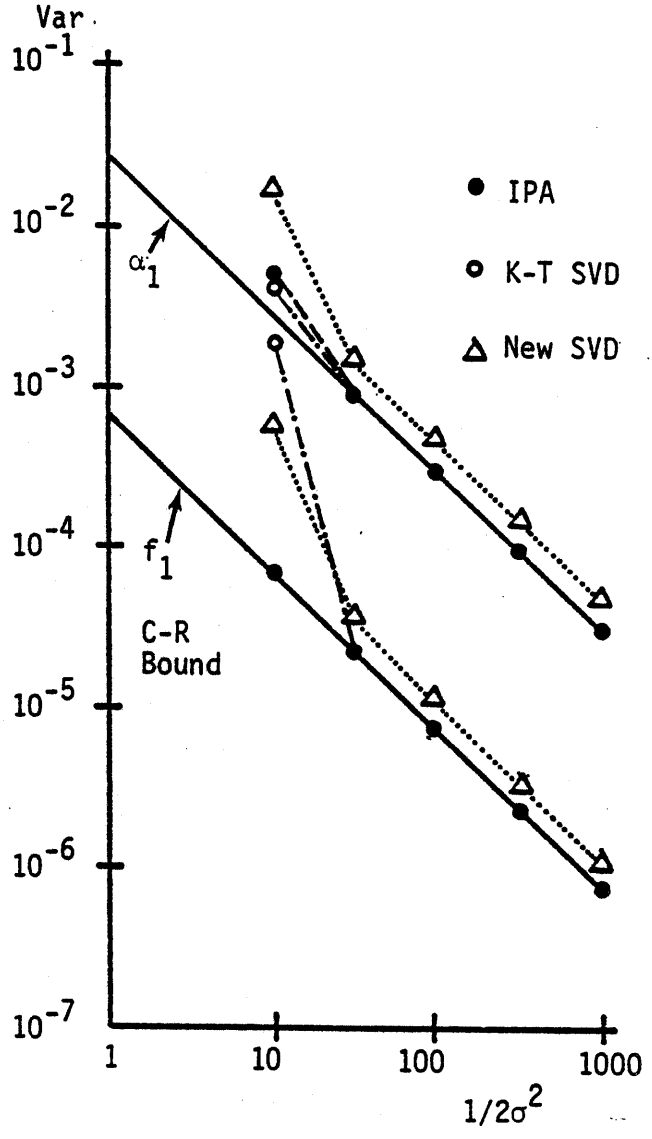


Figure 9. Comparison of pole estimators.

($\alpha_1 = 0.2$, $f_1 = 0.52$)

Table 1. Comparison of SVD Methods

SNR (dB)	$\lvert\hat{\alpha}_1 - \alpha_1\rvert$ New SVD	$\lvert\hat{\alpha}_1 - \alpha_1\rvert_{K\text{-}T}$ Unmodified	$\lvert\hat{\alpha}_1 - \alpha_1\rvert_{K\text{-}T}$ Modified
30	0.7382×10^{-4}	0.3574×10^{-3}	0.3538×10^{-3}
25	0.2873×10^{-3}	0.3059×10^{-2}	0.6849×10^{-3}
20	0.6458×10^{-3}	0.9733×10^{-2}	0.1279×10^{-2}
15	0.1365×10^{-2}	0.3174×10^{-1}	$^{0}.1132\times10^{-2}$

For K = 2, we set $\alpha_1 = 0.1$, $\alpha_2 = 0.2$, $f_1 = 0.52$, and $f_2 = 0.42$. Figure 10 shows that IPA performs well up to 15 of SNR.

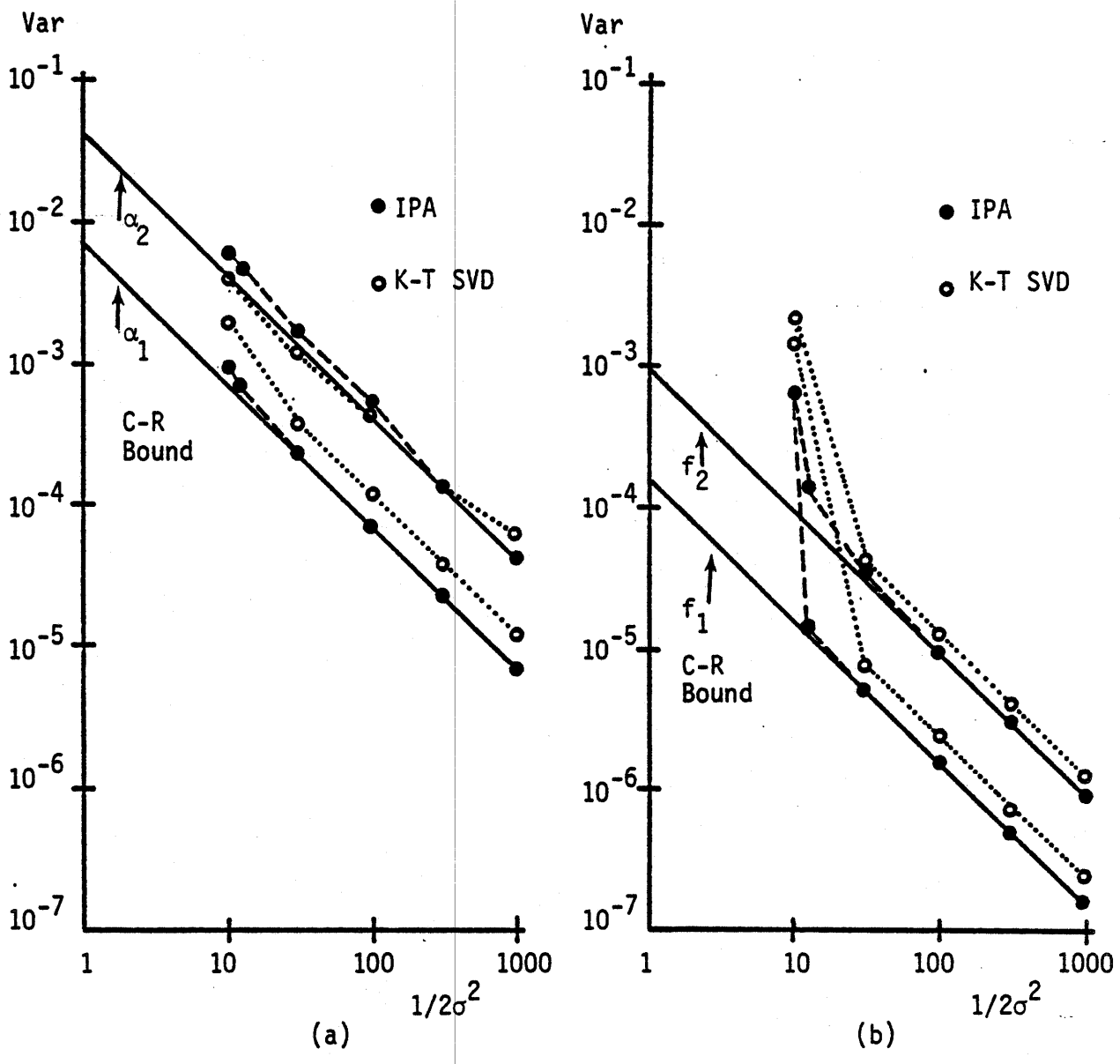


Figure 10. (a),(b) Comparison of pole estimators.
($\alpha_1 = 0.1$, $\alpha_2 = 0.2$, $f_1 = 0.52$, $f_2 = 0.42$)

CHAPTER 6

APPLICATION OF THE IPA TO THE SEM

The singularity expansion method (SEM) form of the surface current density on a finite-dimension, perfectly conducting object in free space as shown in Figure 11 is written as [2,3,4]

$$\vec{J}_s(\vec{r}_s,t)$$

$$= E_o \sum_k \tilde{f}_p(s_k) n_k^{(max)} n_k(\vec{I}_I,\vec{I}_p)\vec{j}_{s_k}(\vec{r}_s)\exp(s_k t)u(t-t_o) \qquad (68)$$

$$+ \text{other SEM terms} + \text{noise}$$

where

t_0 = turn-on time

$\vec{j}_{s_k}(\vec{r}_s)$ = natural mode (appropriately normalized)

s_k = natural frequency

$n_k^{(max)}$ = normalization factor

n_k = coupling coefficient (appropriately normalized)

$\vec{I}_I$ = direction of incidence

$\vec{I}_p$ = polarization vector

$\tilde{f}_p(s)$ = Laplace transform of incident waveform $f_p(t)$

E_0 = scaling constant for incident wave (in V/m)

$\vec{r}_s$ = coordinate on the surface S of the object.

The problem of computing aircraft SEM parameters has been considered by several investigators. Because of the complexity of aircraft surface geometry, it is difficult to formulate and solve the scattering problem

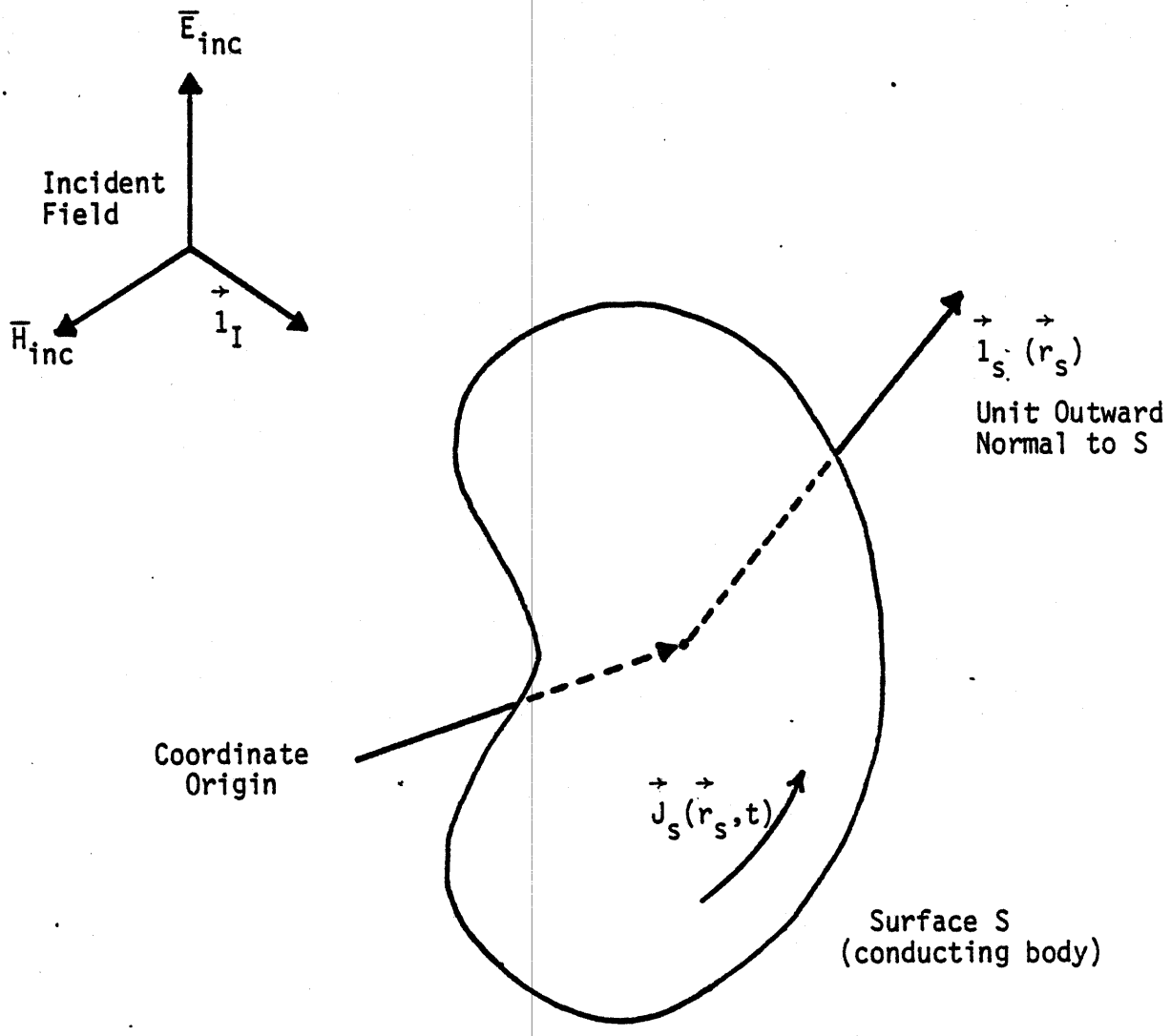


Figure 11. Geometry of the conducting body and incident field.

analytically for an exact model. When an electromagnetic pulse (EMP) strikes an aircraft, it induces a surface current. For EMP testing, several sensors are placed at different locations of the aircraft surface to measure this current. Also the angle between the incidence direction and an aircraft is changed to obtain multiple EMP data sets for several directions of incidence at each observation location. The source of noise is mostly from nonlinear distortion and quantization in the instrumentation system. A typical number of samples for each data set is 512 with a sampling interval of 5 ns. It is desired to estimate SEM parameters of Equation (68) given noisy multiple observations.

As a special case of the general Equation (68), we consider a simple thin wire scatterer. Suppose the wire is struck by a plane electromagnetic wave. The direction of propagation $\vec{I}_I$ of the incident field makes an angle θ with the normal to the center of the scatterer as shown in Figure 12. The induced current on the scatterer for any θ and $0 \leq x \leq X$ is represented as [19].

$$I(t,\theta,\ell) = \sum_k n_k^{(max)} n_k(\theta) i_k(x) \tilde{f}(s_k) \exp(s_k t) u(t-t_o) \qquad (69)$$

$$+ \text{ other SEM terms} + \text{noise}$$

Note that the parameters in (68) are now simplified to a one-dimensional case. Unlike the complicated aircraft case, the SEM parameters for a thin wire have been determined by theoretical methods based on Maxwell's equations [3]. The normalized natural mode, which describes the spatial amplitude variation in the induced current, is given approximately by a simple equation

$$i_k(x) = \sin(k\pi x/X).$$

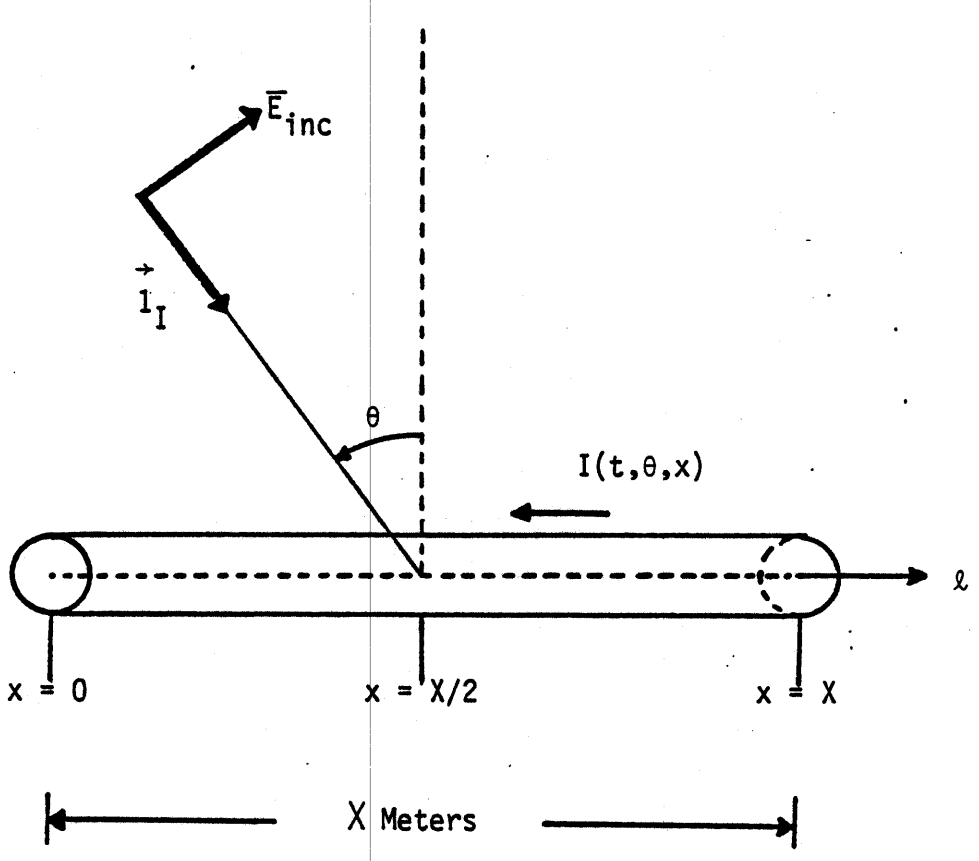


Figure 12. Geometry of the thin-wire scatterer and incident field.

There is no simple equation for the coupling coefficients. But from Figure 6, it is easy to see, for example, that $\eta_k(90°) = 0$ for all k and that $\eta_2(0°) = 0$, etc.

Hereafter we consider an impulse as an exciting function so that $\tilde{f}(s_k) = 1$. We also assume that the other SEM terms are zero, and that the noise is additive Gaussian white noise. In general it is hard to estimate natural frequencies, coupling coefficients, and natural modes at a same time. Our strategy is first to get an estimate of poles using multiple data sets. We discuss this problem in the present chapter. In the next chapter we show how to estimate the other parameters.

6.1 POLE ESTIMATION USING MULTIPLE OBSERVATIONS

In this section, we present an algorithm for pole estimation given multiple observations. Equation (68) shows that measurements made at different locations or with different directions of incidence have the same poles. We can use multiple data sets efficiently to get an improvement in the estimate of natural frequencies. A currently existing simple technique for using multiple data sets is simply to compute poles for each data set and then average the results [20]. This is not the best approach as we argue below.

Let the vector $y^{(\ell,i)}$ be the observation vector at each location x_ℓ ($\ell = 1, 2, \ldots, L$) for each incidence direction θ_i ($i = 1, 2, \ldots, I$). There are $M = LI$ observations available. Without considering incidence directions and observation locations, any observation vector for $m = 1, 2, \ldots, M$ can be represented as

$$y_n^{(m)} = \sum_{k=1}^{K} c_k^{(m)} z_k^n + e_n^{(m)} \qquad (70)$$

for $n = 0, 1, \ldots, N-1$

where $e_n^{(m)}$ is zero-mean, uncorrelated noise. Because natural frequencies do not depend on m, i.e., each data set shares the same transfer function denominator coefficients but has different numerator coefficients, we set up an equation as follows.

$$e^{(m)} = F\tilde{Y}_L^{(m)} x_I + Fy^{(m)} - F_L a^{(m)} \tag{71}$$

for $m = 1, 2, \ldots, M$

where $(e^{(m)})$ are noise vectors and F, $\tilde{Y}_L^{(m)}$, $y^{(m)}$, $a^{(m)}$, x_I are defined in (31). Suppose the noise variance is the same for all m. By introducing the augmented vector $\vec{e}$, (71) becomes

$$\begin{bmatrix} e^{(1)} \\ e^{(2)} \\ \cdot \\ \cdot \\ e^{(M)} \end{bmatrix} = \begin{bmatrix} F & & 0 \\ & F & \\ & \cdot & \\ & & \cdot \\ 0 & & F \end{bmatrix} \left(\begin{bmatrix} \tilde{Y}_L^{(1)} \\ \tilde{Y}_L^{(2)} \\ \cdot \\ \cdot \\ \tilde{Y}_L^{(M)} \end{bmatrix} x_I + \begin{bmatrix} y^{(1)} \\ y^{(2)} \\ \cdot \\ \cdot \\ y^{(M)} \end{bmatrix} \right) - \begin{bmatrix} F_L & & 0 \\ & F_L & \\ & \cdot & \\ & & \cdot \\ 0 & & F_L \end{bmatrix} \begin{bmatrix} a^{(1)} \\ a^{(2)} \\ \cdot \\ \cdot \\ a^{(M)} \end{bmatrix}$$

$$\vec{e} = \vec{F} \left(\vec{Y}_L x_I + \vec{y} \right) - \vec{F}_L \vec{a} \tag{72}$$

Now we want to minimize $J = \vec{e}^{*} \vec{e}$. If the noise is Gaussian then the x_I and $\vec{a}$ that minimize J are the maximum likelihood estimates. Note that J is a sum of squares of all the errors. Fixing $\vec{F}$ as a constant matrix and setting the derivatives of J with respect to x_I and $\vec{a}$, we get two conditions

$$\vec{F}_L^{\,*}[\vec{F}\vec{Y}_L\, x_I + \vec{y} - \vec{F}_L\vec{a}] = 0 \tag{73a}$$

$$(\vec{F}\vec{Y}_L)^*[\vec{F}\vec{Y}_L x_I + \vec{y} - \vec{F}_L\vec{a}] = 0 \tag{73b}$$

It is easy to see that the matrix $\vec{F}_L^{\,*}\vec{F}_L$ is nonsingular. Thus, from (73a) we get

$$\vec{a} = (\vec{F}_L^{\,*}\vec{F}_L)^{-1}\vec{F}_L^{\,*}\vec{F}[\vec{Y}_L x_I + \vec{y}] \tag{74}$$

Substituting (74) into (73b) we obtain

$$(\vec{F}\vec{Y}_L)^*\vec{Q}\vec{F}\vec{Y}_L\, x_I = -(\vec{F}\vec{Y}_L)^*\vec{Q}\vec{F}\vec{y} \tag{75}$$

where $\vec{Q} = I - \vec{F}_L(\vec{F}_L^{\,*}\vec{F}_L)^{-1}\vec{F}_L^{\,*}$

(75) is rewritten simply as

$$\Big[\sum_m [\tilde{FY}_L(m)]^*\tilde{Q}\,\tilde{FY}_L(m)\Big]\, x_I = \Big[\sum_m [\tilde{FY}_L(m)]^*\tilde{Q}\,Fy(m)\Big] \tag{76}$$

where $\tilde{Q} = I - F_L(F_L^{\,*}F_L)^{-1}F_L^{\,*}$

Since F (and so F_L) depends on x_I, we can use (76) but with iterations. It should be noted that the computational burden associated with this algorithm is considerably less than that with the averaging scheme which takes an average of poles obtained using the IPA for each data

set. The latter calculates inverse filter coefficients, and $(F_L^* F_L)^{-1}$ at each iteration, and the roots of the polynomials after obtaining convergence for each data set, but the former calculates these just once.

So far, the assumption was that the variance of the noise is the same for all m. But in general, the noise variance may be different for each data set. If the SNR is known a priori, it is possible to weigh the data sets differently for each m. For example, the weight for the data set, whose SNR is relatively smaller than others, should be smaller than other weights. Because the Equation (76) is in quadratic form one possible form of weight is

$$\omega_m = w^{(m)*} w^{(m)} / (N\sigma_m^2)$$

where σ_m^2 is the variance of the noise for mth data set and $w^{(m)}$ is an observation vector without noise. In practice, noise-free data sets are not available so that y can be substituted for w. The purpose of weighing data sets like above is to emphasize the relatively good data sets, and to weaken the bad data set, for instance from a location at which natural mode is supposed to be pretty small for all k. Considering weight (76) is rewritten as

$$\left[\sum_m \omega_m \left[\tilde{FY}_L^{(m)} \right]^* \tilde{Q}\, \tilde{FY}_L^{(m)} \right] x_I = - \left[\sum_m \omega_m \left[\tilde{FY}_L^{(m)} \right]^* \tilde{Q}\, Fy^{(m)} \right] \tag{77}$$

where $\tilde{Q}$ is as in (76).

Simulation results in Chapter 8 show that both the bias and the sample variance using the Equation (76) are less than those with the averaging scheme.

6.2 C-R BOUND

In this section we calculate the C-R bound for the characteristic equation coefficients, for which sample variances can be compared, given multiple observations. Suppose $(w_n^{(m)})$ are real numbers. Using the results obtained in Chapter 4, it is easy to show that for Gaussian noise the conditional density function of any $y^{(m)}$ is rewritten as

$$p(y^{(m)} \; x_I) = \tag{78}$$

$$(2\pi)^{-N/2}(\det V^{(m)})^{1/2}\exp[-\tfrac{1}{2}(y^{(m)}+\tilde{W}_L^{(m)}x_I)^T(V^{(m)})^{-1}(y^{(m)}+\tilde{W}_L^{(m)}x_I)]$$

$$\text{where } (V^{(m)})^{-1} = \frac{1}{\sigma_m^2}\,F^T[I - F_L(F_L^T F_L)^{-1}F_L^T]\,F = \frac{1}{\sigma_m^2}\,\tilde{P}^{-1}$$

Thus the conditional density function of the augmented vector $\vec{y}$ becomes

$$p(\vec{y} \; x_I) = \tag{79}$$

$$(2\pi)^{-NM/2}(\det\vec{V})^{1/2}\exp[-\tfrac{1}{2}(\vec{y}+\vec{W}_L\,x_I)^T\vec{V}^{-1}(\vec{y}+\vec{W}_L x_I)]$$

$$\text{where } \vec{V}^{-1} = \begin{bmatrix} \frac{1}{\sigma_1^2}\tilde{P}^{-1} & & & 0 \\ & \frac{1}{\sigma_2^2}\tilde{P}^{-1} & & \\ & & \ddots & \\ 0 & & & \frac{1}{\sigma_M^2}\tilde{P}^{-1} \end{bmatrix}$$

and the derivative of the log of this is

$$\frac{\partial}{\partial x_I} \log p(\vec{y}|x_I) = - \vec{W}_L^T \vec{V}^{-1}(\vec{y} + \vec{W}_L x_I) .$$

The second derivative of the log of (79) is given by

$$- \frac{\partial}{\partial x_I} \left[\frac{\partial}{\partial x_I} \log p(\vec{y}|x_I) \right]^T = \vec{W}_L^T \vec{V}^{-1} \vec{W}_L$$

Therefore, C-R bound for the coefficients of the characteristic equation is given by

$$\text{var}\,(\hat{\theta}_i - \theta_i) \geq \left[\left[\sum_m \frac{1}{\sigma_m^2} \tilde{W}_L^T(m) \tilde{P}^{-1} \tilde{W}_L(m) \right]^{-1} \right]_{ii} \qquad (80)$$

where $\theta_i = b_{K-i+1}$ for $i = 1, 2, \ldots, K$.

If the noise variance is the same for all m, then (84) becomes

$$\text{var}\,(\hat{\theta}_i - \theta_i) \geq \sigma^2 \left[\left[\sum_m \tilde{W}_L^T(m) \tilde{P}^{-1} \tilde{W}_L(m) \right]^{-1} \right]_{ii} . \qquad (81)$$

CHAPTER 7

ESTIMATION OF COUPLING COEFFICIENTS AND NATURAL MODES

In this chapter an algorithm for estimating coupling coefficients, natural modes, and normalization factors given natural frequencies is presented.

The induced current data on the scatterer for any incidence direction θ_i and observation location x_ℓ is represented as (from (69))

$$y_n^{(i,\ell)} = \sum_{k=1}^{K} \eta_k^{(max)} \eta_k^{(i)} j_k^{(\ell)} z_k^n + e_n^{(i,\ell)} \qquad (82)$$

for $n = 0, 1, \ldots, N-1$

where $\eta_k^{(max)}$ is the normalization factor, $\eta_k^{(i)}$ is the normalized coupling coefficients, $j_k^{(\ell)}$ is the normalized natural mode, and $e_n^{(i,\ell)}$ is the noise which is assumed to be zero-mean and uncorrelated. Note that

$$\max\left(\eta_k^{(i)}\right) = 1, \text{ and } \max\left(j_k^{(\ell)}\right) = 1$$

for each k. Let

$$H = \text{diag}\left[\eta_1^{(max)}, \eta_2^{(max)}, \ldots, \eta_K^{(max)}\right],$$

$$G^{(i)} = \text{diag}\left[\eta_1^{(i)}, \eta_2^{(i)}, \ldots, \eta_K^{(i)}\right],$$

$$g^{(i)} = \left[\eta_1^{(i)}, \eta_2^{(i)}, \ldots, \eta_K^{(i)}\right]^T,$$

$$J^{(\ell)} = \text{diag}\left[j_1^{(i)}, j_2^{(\ell)}, \ldots, j_K^{(\ell)}\right], \text{ and}$$

$$j^{(\ell)} = \left[j_1^{(\ell)}, j_2^{(\ell)}, \ldots, j_K^{(\ell)}\right]^T.$$

Also let

$$Z = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ z_1 & z_2 & \cdots & z_K \\ z_1^2 & z_2^2 & \cdots & z_K^2 \\ \vdots & \vdots & & \vdots \\ z_1^{N-1} & z_2^{N-1} & \cdots & z_K^{N-1} \end{bmatrix}.$$

Then in matrix form (82) becomes either

$$y^{(i,\ell)} = Z\, J^{(\ell)}\, \hat{g}^{(i)} + e^{(i,\ell)} \quad \text{or} \tag{83a}$$

$$y^{(i,\ell)} = Z\, G^{(i)}\, \hat{j}^{(\ell)} + e^{(i,\ell)} \tag{83b}$$

where $\hat{g}^{(i)} = Hg^{(i)}$ and $\hat{j}^{(\ell)} = Hj^{(\ell)}$.

Suppose the (z_K) are known. We want to minimize $\vec{e}^{*}\vec{e}$ where $\vec{e}$ is as defined as in (72). Minimizing $\vec{e}^{*}\vec{e}$ with respect to $(\hat{g}^{(i)})$ and $(\hat{j}^{(\ell)})$ is a nonlinear problem even though (z_k) are known a priori. We use an iterative scheme to solve this problem. From (83a)

$$\vec{e}^{*}\vec{e} = \sum_i \sum_\ell [y^{(i,\ell)} - ZJ^{(\ell)}\hat{g}^{(i)}]^{*}[Y^{(i,\ell)} - ZJ^{(\ell)}\hat{g}^{(i)}] \tag{84}$$

Fixing $J^{(\ell)}$ as a constant matrix for all ℓ and setting the derivatives of $\vec{e}^{*}\vec{e}$ with respect to $\hat{g}^{(i)}$ to zero, we obtain

$$\left[\sum_{\ell} (ZJ^{(\ell)})^{*} ZJ^{(\ell)} \right] \hat{g}^{(i)} = \left[\sum_{\ell} (ZJ^{(\ell)})^{*} y^{(i,\ell)} \right] \tag{85}$$

for $i = 1, 2, \ldots, I$.

Similarly using the Equation (83b) $\vec{e}^{*}\vec{e}$ is minimized with respect to $j^{(\ell)}$ by solving following equation

$$\left[\sum_{i} (ZG^{(i)})^{*} ZG^{(i)} \right] \hat{j}^{(\ell)} = \left[\sum_{i} (ZG^{(i)})^{*} y^{(i,\ell)} \right] \tag{86}$$

for $\ell = 1, 2, \ldots, L$.

Let matrices

$$\tilde{G} = [g^{(1)}, g^{(2)}, \ldots, g^{(I)}] ,$$

$$\tilde{J} = [j^{(1)}, j^{(2)}, \ldots, j^{(L)}] ,$$

$$\hat{\tilde{G}} = [\hat{g}^{(1)}, \hat{g}^{(2)}, \ldots, \hat{g}^{(I)}] , \text{ and}$$

$$\hat{\tilde{J}} = [\hat{j}^{(1)}, \hat{j}^{(2)}, \ldots, \hat{j}^{(L)}] .$$

Then from (83a) and (83b),

$$\hat{\tilde{G}} = H\tilde{G} \quad \text{and} \tag{87a}$$

$$\hat{\tilde{J}} = H\tilde{J}. \tag{87b}$$

Note that the normalized coupling coefficient (natural mode) is obtained by normalizing the rows of $\hat{\tilde{G}}$ $(\hat{\tilde{J}})$ so that the maximum element of each row of $\tilde{G}$ $(\tilde{J})$ is 1.

Now the algorithm is as follows.

1) Set $(G^{(i)})$ as the identity matrix, solve (86), and normalize the result to obtain an initial estimate of $(j^{(\ell)})$.

2) Set $(J^{(\ell)})$ using the preceding estimate of $(j^{(\ell)})$ and solve the Equation (85) to obtain $(\hat{g}^{(i)})$.

3) Form $\hat{\tilde{G}}$ and normalize every row of $\hat{\tilde{G}}$ to obtain $\tilde{G}$ (i.e., $g^{(i)}$) and normalization factors $(n_k^{(max)})$.

4) Set $(G^{(i)})$ using the preceding estimate of $(g^{(i)})$ and solve the Equation (86) to obtain $(\hat{j}^{(\ell)})$.

5) Form $\hat{\tilde{J}}$ and normalize every row of $\hat{\tilde{J}}$ to obtain $\tilde{J}$ (i.e., $J^{(\ell)}$) and the normalization factors $(n_k^{(max)})$.

6) If convergence is obtained then stop. If not then go to 2) and continue.

If convergence is obtained, then $\vec{e}^{*}\vec{e}$ given (z_k) is minimized. The normalization factors obtained both in procedure 3) and 5) may be averaged to obtain a better estimate.

CHAPTER 8

SIMULATION RESULTS

Experiments I and II in this chapter are to support the algorithms proposed in Chapters 6 and 7 respectively.

8.1 EXPERIMENT I

Multiple data sets which share the same poles but have different residues were considered in this experiment. The system order was chosen to be K = 2. Data sets consisting of a pair of complex conjugate damped sinusoids in white noise were used. The data set for any m (m = 1, ..., M) is represented as

$$y_n^{(m)} = c_1^{(m)} z_1^n + \bar{c}_1^{(m)} \bar{z}_1^n + e_n^{(m)}$$

Number of data points N was 20 and the standard deviation of the noise was 0.1 for all m.

For the first example, four data sets were used. The pole z_1 was -0.7 + j0.5 so that the characteristic equation

$$B_1(z) = 1 + 1.4z^{-1} + 0.74z^{-2}$$

The residues were: $c_1^{(1)} = 1$, $c_1^{(2)} = 0.3 + j0.4$,

$c_1^{(3)} = -0.4 + j0.2$, $c_1^{(4)} = 0.7 + j0.7$.

Method I is the averaging scheme which takes the average of coefficients obtained using the IPA for each data set separately. Method II is the algorithm presented in Chapter 6. The C-R bound for the coefficients was calculated using noise free data. The variances with respect to the true coefficients were calculated after 100 trials for each method.

Table 2(a) shows that both the bias and the sample variance with method II were about the half of those with method I. Note that sample variances with method II were very close to the C-R bounds.

For the second example, 6 data sets were used and $z_1 = -0.8 + j0.4$ so that

$$B_2(z) = 1 + 1.6z^{-1} + 0.8z^{-2} .$$

The residues were: $c_1^{(1)} = 1$, $c_1^{(2)} = 0.3 + j0.5$

$c_1^{(3)} = -0.5 + j0.2$, $c_1^{(4)} = 0.8 + j0.3$,

$c_1^{(5)} = -0.7 + j0.7$, $c_1^{(6)} = 0.6 + j0.2$.

Table 2(b) shows that method II outperformed method I.

For the third example, 10 data sets were used and $z_1 = 0.75 + j0.48$ so that

$$B_3(z) = 1 + 1.5z^{-1} + 0.7929z^{-2} .$$

The residues were: $c_1^{(1)} = 1$, $c_1^{(2)} = 0.8 + j0.2$,

$c_1^{(3)} = 0.7 + j0.4$, $c_1^{(4)} = -0.5 + j0.3$,

$c_1^{(5)} = -0.4 + j0.3$, $c_1^{(6)} = 0.2 + j0.5$,

$c_1^{(7)} = 0.3 + j0.3$, $c_1^{(8)} = -0.2 + j0.3$,

$c_1^{(9)} = j1$, $c_1^{(10)} = 0.2 + j0.8$.

Table 2. Comparison of Pole Estimators

	bias: $\lvert\hat{b}_i - b_i\rvert$		var $(\hat{b}_i - b_i)$		
	I	II	I	II	C-R Bound
b_1	8.24×10^{-3}	3.76×10^{-3}	3.85×10^{-4}	1.67×10^{-4}	1.79×10^{-4}
b_2	5.88×10^{-3}	2.89×10^{-3}	2.67×10^{-4}	1.08×10^{-4}	1.12×10^{-4}

(a)

	bias: $\lvert\hat{b}_i - b_i\rvert$		var $(\hat{b}_i - b_i)$		
	I	II	I	II	C-R Bound
b_1	4.50×10^{-3}	3.49×10^{-3}	1.01×10^{-4}	6.16×10^{-5}	5.54×10^{-5}
b_2	3.39×10^{-3}	2.54×10^{-3}	7.74×10^{-5}	4.99×10^{-5}	4.42×10^{-5}

(b)

	bias: $\lvert\hat{b}_i - b_i\rvert$		var $(\hat{b}_i - b_i)$		
	I	II	I	II	C-R Bound
b_1	4.97×10^{-3}	7.46×10^{-4}	8.53×10^{-5}	5.22×10^{-5}	3.94×10^{-5}
b_2	4.62×10^{-3}	1.30×10^{-3}	8.11×10^{-5}	3.95×10^{-5}	3.38×10^{-5}

(c)

Again Table 2(c) shows that method II was a better estimator than method I. Especially the bias for b_1 with method II is about the one-fifth of that with method I.

From all three examples, sample variances were pretty close to the C-R bounds.

8.2 EXPERIMENT II

In this experiment we used 25 data sets from 5 observation locations for 5 incidence directions. The SEM parameter values are representative of those for a thin wire [19]. Any data set for $i = 1, \ldots, 5$ and $\ell = 1, \ldots, 5$ is given by

$$y_n^{(i,\ell)} = \sum_{k=1}^{3} \left(\eta_k^{(\max)} \eta_k^{(i)} j_k^{(\ell)} z_k^n + \bar{\eta}_k^{(\max)} \bar{\eta}_k^{(i)} \bar{j}_k^{(\ell)} \bar{z}_k^n \right) + e^{(i,\ell)}$$

for $n = 0, 1, \ldots, 49$

where $e^{(i,\ell)}$ is an uncorrelated zero-mean noise vector with standard deviation σ.

Three complex conjugate z-pole pairs (s-poles in [20] were translated) were:

$z_1 = 0.5589 + j0.7325$, $z_2 = -0.2831 + j0.8396$,

and $z_3 = -0.8320 + j0.2237$

Normalization factors were:

$\eta_1^{(\max)} = -7.0728 - j1.9371$, $\eta_2^{(\max)} = 1.4320 - j4.6624$,

and $\eta_3^{(\max)} = 3.9331 + j1.2755$

Actual normalization factors in [20] were the above numbers multiplied by 10^5.

Coupling coefficients were:

	θ_1	θ_2	θ_3	θ_4	θ_5
k = 1	1.0	0.92-j0.005	0.76-j0.008	0.60-j0.012	0.28-j0.01
k = 2	0.02	0.76+j0.01	1.0	0.92-j0.021	0.48-j0.028
k = 3	-0.74-j0.018	j0.02	0.76+j0.025	1.0	0.60-j0.03

where $\theta_1 = 0°$, $\theta_2 = 20°$, $\theta_3 = 35°$, $\theta_4 = 48°$, and $\theta_5 = 70°$.

Natural modes were taken as:

	x_1	x_2	x_3	x_4	x_5
k = 1	0.52-j0.01	0.71-j0.007	1.0	0.71-j0.007	0.52-j0.01
k = 2	0.088-j0.014	1.0	0.04	-1.0	-0.88+j0.014
k = 3	1.0	0.71+j0.032	-0.96	0.72+j0.032	1.0

where five locations on the thin wire were at 0.16 m, 0.245 m, 0.5 m, 0.755 m, and 0.84 m (the length of the wire was 1 m). The plot of normalized coupling coefficients and natural modes are shown in Figures 13-16.

For the first example, $\sigma = 0.5$. Exact values of z-poles were given. Each parameter converged to within an order of 10^{-6} at 5th iteration. The sum of the absolute values of the estimated parameter errors is shown in Table 3 as a measure of fit.

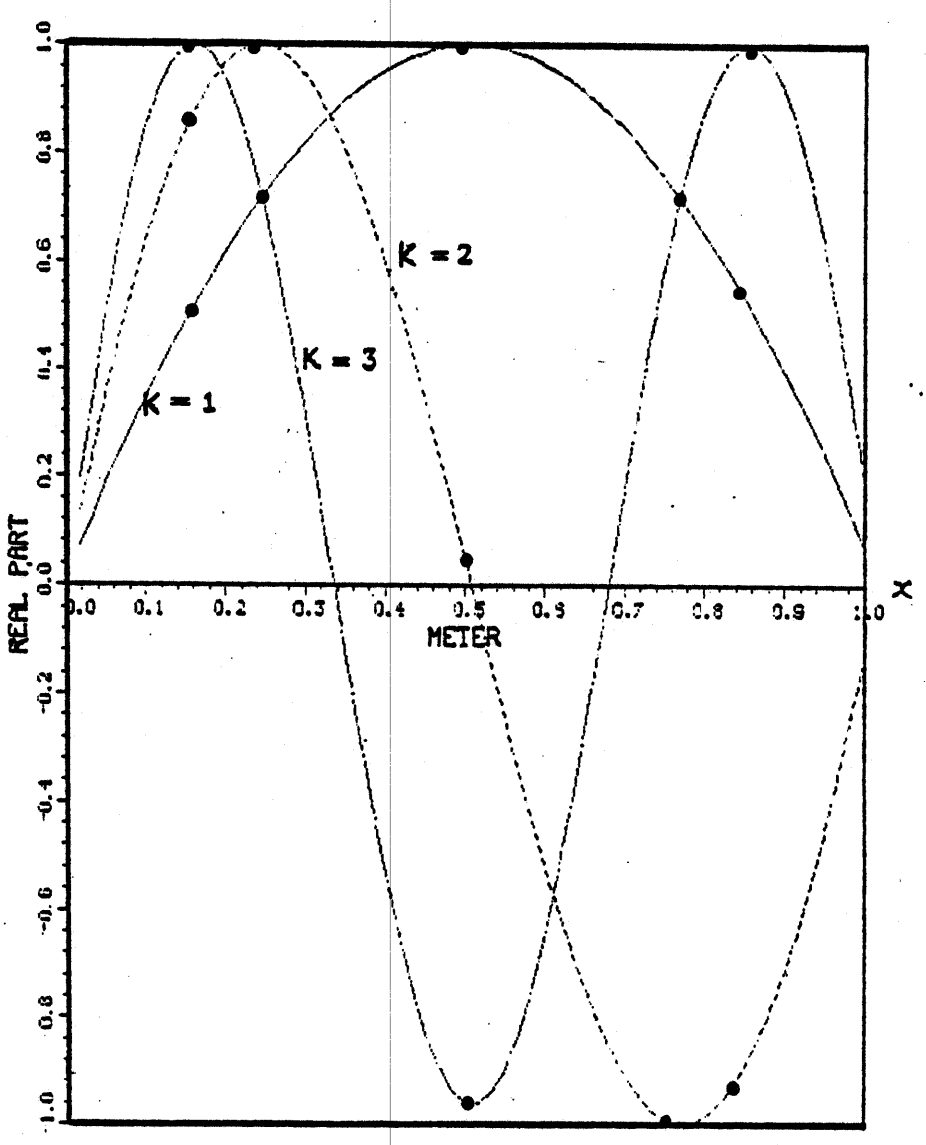


Figure 13. Real parts of first three normalized natural modes for thin wire.

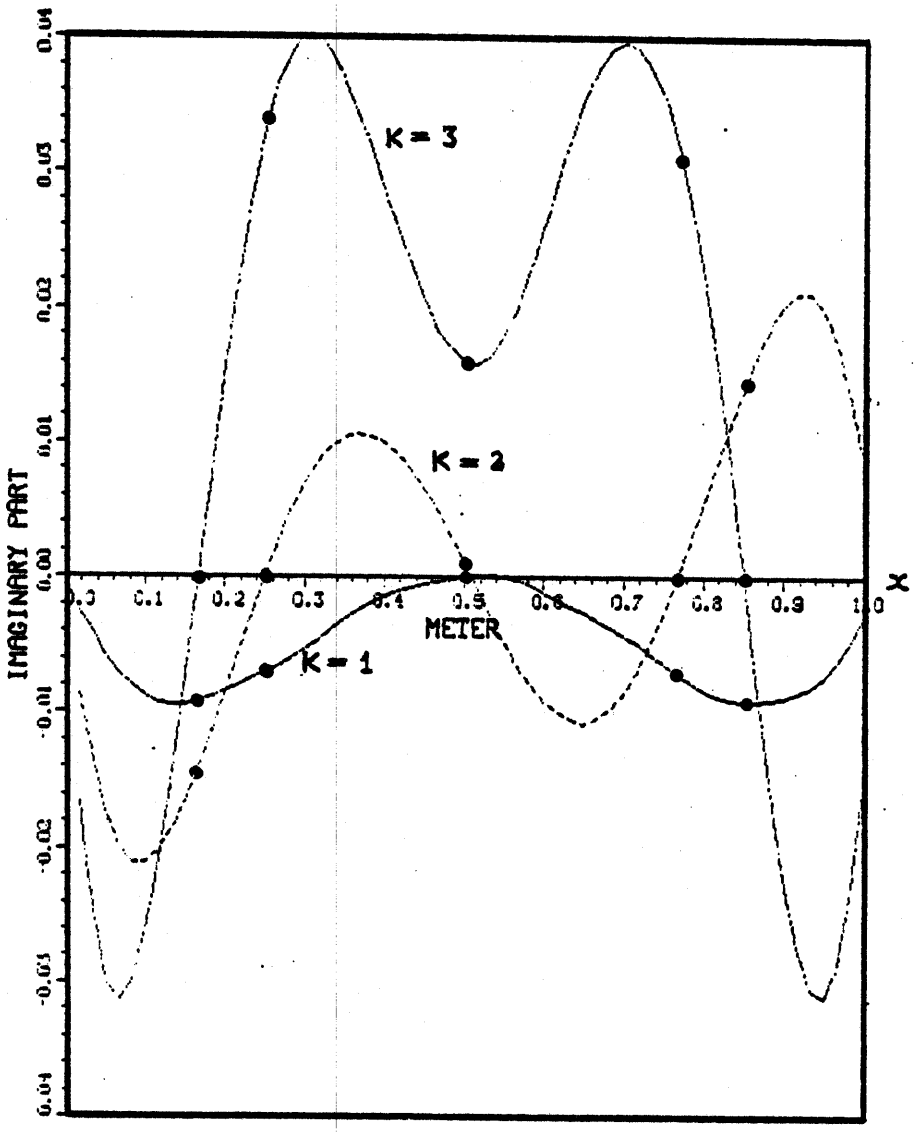


Figure 14. Imaginary parts of first three normalized natural modes for thin wire.

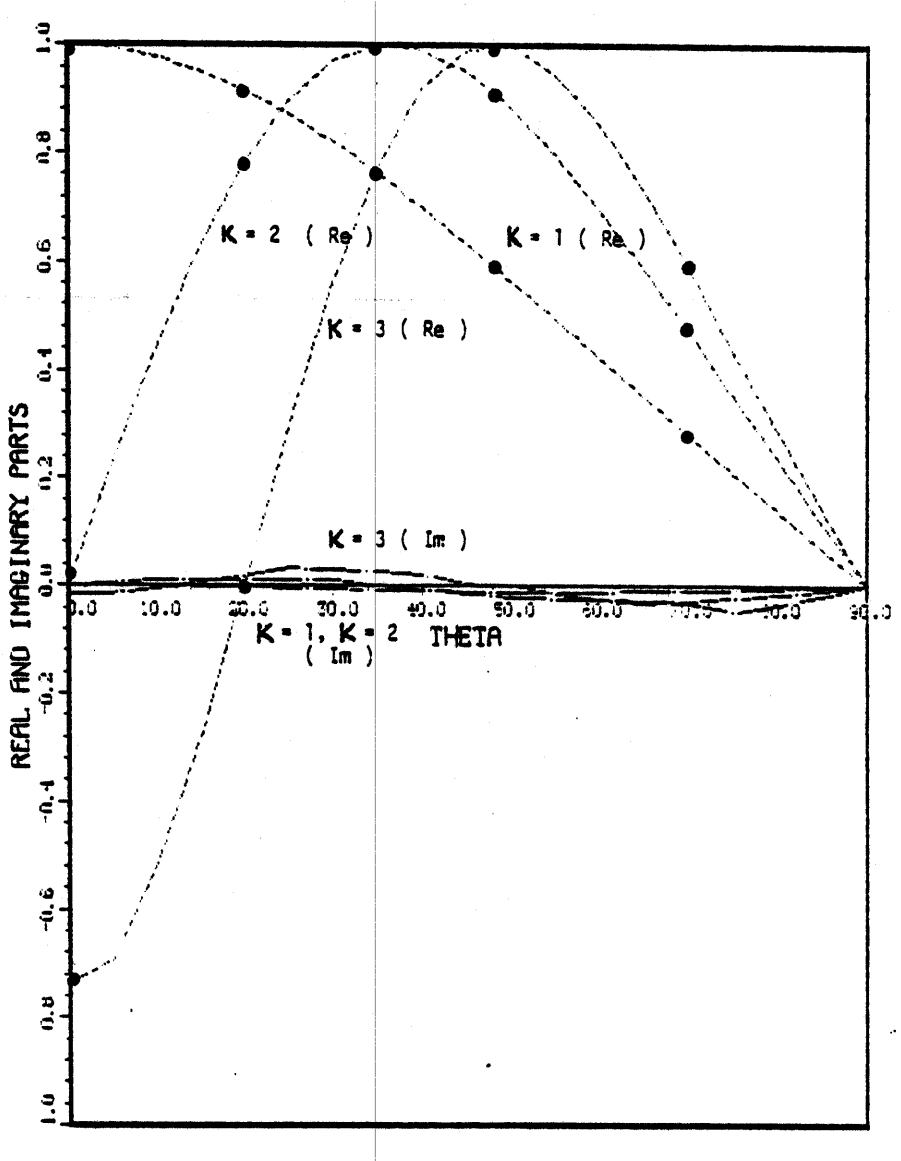


Figure 15. Normalized coupling coefficients calculated using Gaussian excitation, both real and imaginary parts of first three modes, are shown.

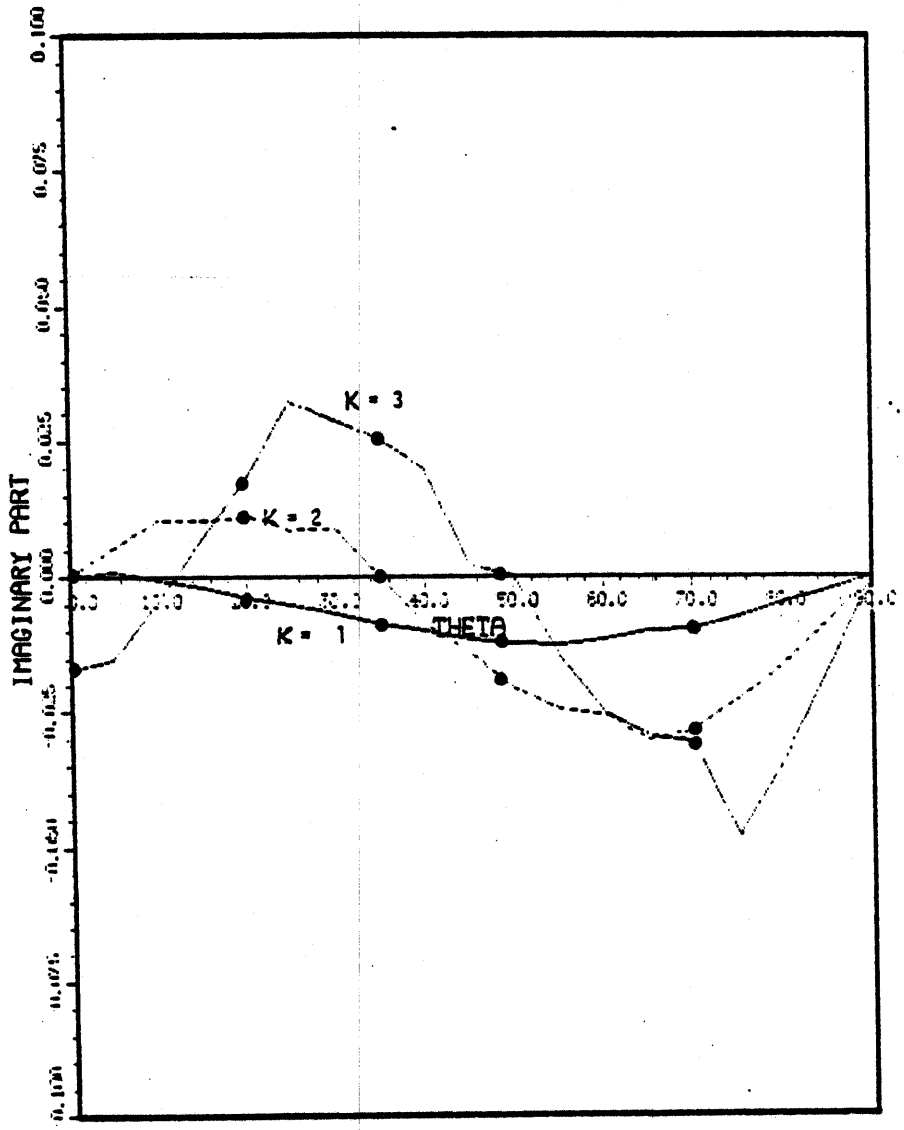


Figure 16. Imaginary parts of first three modes of normalized coupling coefficients (Gaussian excitation).

We intentionally omit the results of some iterations because there were no big changes in the errors after the 2nd iteration. Table 3 shows that the errors in the natural modes decreased significantly at the 1st iteration. But after the 1st iteration the errors seem to remain almost the same.

Table 3. Sum of Errors at Each Iteration (σ = 0.5)

	$\sum_k \sum_i \left\|\hat{n}_k^{(i)} - n_k^{(i)}\right\|$	$\sum_k \sum_\ell \left\|\hat{j}_k^{(\ell)} - j_k^{(\ell)}\right\|$	$\sum_k \left\|\hat{n}_k^{(max)} - n_k^{(max)}\right\|$
init. est.	---	0.74923	---
1st iter.	0.23367	0.41741	0.19006
2nd iter.	0.23869	0.41665	0.18112
5th iter.	0.23862	0.41653	0.18026

For the second example, σ = 1.0. Again the exact values of the z-poles were given. Convergence was obtained at the 6th iteration. Table 4 shows that the errors in the natural modes decreased by about half of at the 1st iteration. For both examples, errors in the normalization factors decreased at each iteration. We performed some more experiments for $\sigma < 0.5$. In all cases, the errors in the natural modes decreased significantly only at the 1st iteration.

Table 4. Sum of Errors at Each Iteration (σ = 1.0)

	$\sum_k \sum_i \left\lvert \hat{\eta}_k^{(i)} - \eta_k^{(i)} \right\rvert$	$\sum_k \sum_\ell \left\lvert \hat{j}_k^{(\ell)} - j_k^{(\ell)} \right\rvert$	$\sum_k \left\lvert \hat{\eta}_k^{(max)} - \eta_k^{(max)} \right\rvert$
init. est.	---	1.43337	---
1st iter.	0.69081	0.71039	1.36592
2nd iter.	0.62938	0.72258	1.30901
6th iter.	0.62042	0.72486	1.30044

CHAPTER 9

CONCLUSIONS

In this dissertation we have introduced two new algorithms for estimating poles given noise-contaminated impulse response data. One is an algorithm to extract the reduced characteristic equation from the weakest eigenvectors when the system order is overdetermined. This algorithm is closely related to Henderson's deflation algorithm [16] and simulation results show that new method performed better than the existing method. The other algorithm is the iterative preprocessing algorithm (IPA) which is related to the Steiglitz-McBride method [5,7] and to the Evans-Fischl technique [18] but has an advantage over either in both stability and computational complexity. The approximate IPA (AIPA) which reduces the computational burden further was described. The AIPA for the pure sinusoid case is related to Kay's IFA [7].

Also the C-R bound for the estimated characteristic equation coefficients, which is a valuable tool for evaluating different estimators without finding roots of the equation was evaluated.

The IPA was extended to SEM parameter estimation. Using the IPA, it is possible to process multiple data sets at the same time to get an improvement in pole estimation. Finally, an iterative scheme to estimate coupling coefficients, and natural modes was introduced. Simulation results indicate that the estimation errors decrease most at the first iteration after the initial estimate. Together these results provide a way to estimate the SEM description of a scatterer from multiple data sets taken at different locations and with different directions of

incidence. For large to moderate signal-to-noise ratios these estimates meet the C-R bound and thus have the minimum variance of any unbiased estimators.

So far the theoretical convergence and accuracy properties of the IPA are not known. The IPA assumes the system order is given. The next phase of research should be to investigate the convergence property of the IPA, and to devise an algorithm combined with order selection. Another interesting problem would be to improve the AIPA. As discussed in Chapter 3, the AIPA approximates $(F_L^* F_L)^{-1}$. It would be useful to have an approximation for $F_L(F_L^* F_L)^{-1} F_L^*$ instead.

APPENDIX A

The data matrix W and the matrix G are as in (13) and (15) respectively.

Th. Every column g_m of G is in the null space of W.

(proof) Let w_j be the jth row of the matrix W. If $w_j g_m = 0$ for all $j = 1, \ldots, N-M$ and $m = 1, \ldots, M-K+1$, then the proof is done. Let $= [1, z, \ldots, z^M]^T$ and $u_i = [1, z_i, \ldots, z_i^M]^T$ where (z_i) are system poles. And let $v = [1, z, \ldots, z^K]^T$, $v_i = [1, z_i, \ldots, z_i^K]^T$ and $x = [b_K, b_{K-1}, \ldots, b_0]^T$. Then $g_m^T u_i = z_i^{m-1}(x^T V_i) = 0$ for all m,i.

Now

$$w_j = [\sum_{i=1}^{K} c_i z_i^{j-1}, \sum_{i=1}^{K} c_i z_i^{j}, \ldots, \sum_{i=1}^{K} c_i z_i^{j+M-1}]$$

$$= [\sum_{i=1}^{K} c_i z_i^{j-1}(1), \sum_{i=1}^{K} c_i z^{j-1}(z_i), \ldots, \sum_{i=1}^{K} c_i z_i^{j-1}(z_i^M)]$$

$$= \sum_{i=1}^{K} \beta_i u_i^T \quad (\beta_i = c_i z^{j-1})$$

Thus,

$$w_j g_m = g_m^T w_j^T = g_m^T \sum_{i=1}^{K} \beta_i u_i = \sum_{i=1}^{K} \beta_i g_m^T u_i = 0$$

Q.E.D.

APPENDIX B

Here we want to show that the Equation (36) is identical to the Equation (40).

By partitioning, F and $\tilde{Y}$ become

$$F = \underbrace{[F_L}_{K} \underbrace{F_R]}_{N-L} \qquad F_L = \begin{bmatrix} F_0 \\ \hline F_1 \end{bmatrix} \begin{matrix} \} K \\ \} N-K \end{matrix} \qquad F_R = \begin{bmatrix} 0 \\ \hline F_2 \end{bmatrix} \begin{matrix} \} K \\ \} N-K \end{matrix}$$

(B-1)

$$\tilde{Y} = \begin{bmatrix} Y_0 \\ \hline Y \end{bmatrix} \begin{matrix} \} K \\ \} N-K \end{matrix} = \begin{bmatrix} Y_{OL} & y_{OR} \\ \hline \underbrace{Y_L}_{K} & \underbrace{y_R}_{1} \end{bmatrix}$$

Using (B-1), (36) becomes

$$(B\text{-}2) \qquad (F_L Y_{OL} + F_R Y_L)^* \tilde{Q} (F_L Y_{OL} + F_R Y_L) x_I$$

$$= - (F_L Y_{OL} + F_R Y_L)^* \tilde{Q} (F_L y_{OR} + F_R y_R)$$

$$\text{where } \tilde{Q} = I - F_L (F_L{}^* F_L)^{-1} F_L{}^*$$

Equation (B-2) is simplified to

$$(B\text{-}3) \qquad (F_R Y_L)^* [I - F_L (F_L{}^* F_L)^{-1} F_L{}^*] (F_R Y_L) x_I$$

$$= - (F_R Y_L)^* [I - F_L (F_L{}^* F_L)^{-1} F_L{}^*] F_R y_R$$

From (B-1), $F_R^* F_L = F_2^* F_1$ and $F_R^* F_R = F_2^* F_2$.

Thus, (B-3) is rewritten as

$$(B\text{-}4) \qquad (F_2 Y_L)^* [I - F_1 (F_L^* F_L)^{-1} F_1^*] (F_2 Y_L)\, x_I$$

$$= - (F_2 Y_L)^* [I - F_1 (F_L^* F_L)^{-1} F_1^*]\, F_2 y_R$$

Equation (40) is

$$(B\text{-}5) \qquad Y_L^* (DD^*)^{-1} Y_L\, x_I = - Y_L^* (DD^*)^{-1} y_R$$

If $(DD^*)^{-1} = F_2^* [I - F_1 (F_L^* F_L)^{-1} F_1^*]\, F_2$, then (B-4) and (B-5) are identical and the proof is done.

<u>Th.</u> $(DD^*)^{-1} = F_2^* [I - F_1 (F_L^* F_L)^{-1} F_1^*]\, F_2$

(proof) By partitioning, B becomes

$$B = \begin{bmatrix} \overbrace{D_0}^{K} & \overbrace{0}^{N-K} \\ D_1 & D_2 \end{bmatrix} \begin{matrix} \} K \\ \} N-K \end{matrix} \qquad D = [D_1 \mid D_2]$$

It is easily shown that

$$D_0 = F_0^{-1}, \quad D_2 = F_2^{-1}, \quad D_1 = - F_2^{-1} F_1 F_0^{-1}$$

left side $= (DD^*)^{-1} = [D_1D_1^* + D_2D_2^*]^{-1}$

$= [F_2^{-1}\{F_1(F_0^*F_0)^{-1}F_1^* + I\}F_2^{-*}]^{-1} = P$

right side $= F_2^*[I - F_1(F_L^*F_L)^{-1}F_1^*]F_2$

$= F_2^*[I - F_1(F_0^*F_0+F_1^*F_1)^{-1}F_1^*]F_2 = Q$

We want to show that $P^{-1}Q = I$.

Because

$$I + (F_0^*F_0)^{-1}F_1^*F_1 = I + (F_0^*F_0)^{-1}F_1^*F_1,$$

it is obvious that

$$(F_0^*F_0)^{-1}(F_0^*F_0 + F_1^*F_1) - I = (F_0^*F_0)^{-1}F_1^*F_1.$$

Postmultiplying $(F_0^*F_0 + F_1^*F_1)^{-1}$ on both sides yields

$$(F_0^*F_0)^{-1} - (F_0^*F_0 + F_1^*F_1)^{-1}$$

$$= (F_0^*F_0)^{-1}F_1^*F_1(F_0^*F_0 + F_1^*F_1)^{-1}.$$

Thus, the following holds

(B-6) $$F_1[(F_0^*F_0)^{-1} - (F_L^*F_L)^{-1} - (F_0^*F_0)^{-1}F_1^*F_1(F_L^*F_L)^{-1}]F_1^*$$

$$= \tilde{T} = 0$$

Finally,

$$P^{-1}Q = F_2^{-1}\,[I + F_1F_0^{-1}F_0^{-*}F_1^{*}]\,F_2^{-*}[I - F_1(F_L^{*}F_L)^{-1}F_1^{*}]\,F_2$$

$$= F_2^{-1}[I + \tilde{T}]\,F_2$$

$$= I \quad \text{(by B-6)}$$

Q.E.D.

APPENDIX C

Let F(N) be an NxN matrix as defined in Chapter 3. In this appendix, we want to describe what happens to $F_L(N)^* F_L(N)$ for large N. Note that notations are slightly changed in this appendix.

Consider the infinite matrices

$$B = \begin{bmatrix} B_0 & 0 \\ B_3 & B_2 \end{bmatrix} \begin{matrix} \} K \\ \} \infty \end{matrix} \qquad F = B^{-1} = \begin{bmatrix} F_0 & 0 \\ F_1 & F_2 \end{bmatrix} \begin{matrix} \} K \\ \} \infty \end{matrix} \qquad B_3 = \begin{bmatrix} B_1 \\ 0 \end{bmatrix} \begin{matrix} \} K \\ \} \infty \end{matrix}$$

(K, ∞ under B; K, ∞ under F)

It is easy to show that

$$F_0 = B_0^{-1}$$

$$F_2 = B_2^{-1} \tag{C-1}$$

$$F_1 = - F_2 B_3 F_0$$

For convenience, assume matrices B and F are real. But the result can be extended to complex case immediately. Also assume $\sum_{j=i}^{\infty} f_j f_{j-i} < \infty$ for $i = 0, 1, \ldots, K$.

It is obvious that

$$R = \lim_{N \to \infty} F_L^T(N) F_L(N) = F_0^T F_0 + F_1^T F_1$$

R can be interpreted as an autocorrelation matrix of the inverse filter coefficients such that

$$(C\text{-}2) \quad R = \begin{bmatrix} r_0 & r_1 & \cdots & r_{K-1} \\ r_1 & r_0 & \cdots & r_{K-2} \\ \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & & \cdot \\ r_{K-1} & r_{K-2} & \cdots & r_0 \end{bmatrix} \qquad \text{where } r_i = \sum_{j=i}^{\infty} f_j f_{j-i} \quad \text{and}$$

$$(C\text{-}3) \quad R\, x_I = -\, r \qquad \text{where } r = [r_1, r_2, \ldots, r_K]^T .$$

Because we consider infinite matrices, the following holds (by C-1)

$$F_1 = -\, F_2 B_3 F_0 = -\begin{bmatrix} F_0 & 0 \\ F_1 & F_2 \end{bmatrix} \begin{bmatrix} B_1 \\ 0 \end{bmatrix} F_0 = -\begin{bmatrix} F_0 B_1 F_0 \\ F_1 B_1 F_0 \end{bmatrix}$$

so that

$$(C\text{-}4) \quad F_1^T F_1 = F_0^T B_1^T R\, B_1 F_0$$

Now we want to show that $R^{-1} = B_0 B_0^T - B_1^T B_1$.

<u>Th.</u> $\quad R^{-1} = B_0 B_0^T - B_1^T B_1$

(proof) Denote

$$J = \begin{bmatrix} 0 & & & 1 \\ & & 1 & \\ & \cdot & & \\ & \cdot & & \\ 1 & & & 0 \end{bmatrix}$$

Premultiplying (postmultiplying) a matrix by J results in swapping all the rows (columns) of the original matrix. Then one can show that (using C-3)

$$B_1 R B_0 J$$

is always symmetric. Thus,

(C-5) $B_1^T (R_1 R B_0 J) B_1$ is symmetric.

It can be shown easily that

(C-6a) $JB_1 = B_1^T J$ and that

(C-6b) $B_0 B_1^T J$ is symmetric.

From (C-5),

(C-7) $B_1^T B_1 R B_0 J B_1 = B_1^T B_1 R B_0 B_1^T J$ by (C-6a)

$= (B_0 B_1^T J)^T R B_1^T B_1$ by symmetry

$= B_0 B_1^T J R B_1^T B_1$ by (C-6b)

Because B_1 is nonsingular for $b_K \neq 0$, postmultiplying B_1^{-1} on both sides of (C-7) yields

$$B_1^T B_1 R B_0 J = B_0 B_1^T J R B_1^T$$

$$B_1^T B_1 R B_0 = B_0 B_1^T J R B_1^T J$$

$= B_0 B_1^T (JRJ) B_1$ by (C-6a)

Thus,

(C-8) $$B_1^T B_1 R B_0 = B_0 B_1^T R B_1$$

To show $R^{-1} = B_0 B_0^T - B_1^T B_1$, we need to show

$$(B_0 B_0^T - B_1^T B_1)\, R = I.$$

Substituting for $R = F_0^T F_0 + F_1^T F_1$,

$$(B_0 B_0^T - B_1^T B_1)(F_0^T F_0 + F_1^T F_1)$$

$$= B_0 B_0^T F_1^T F_1 - B_1^T B_1 (F_0^T F_0 + F_1^T F_1) + I$$

$$= B_0 B_0^T (F_0^T B_1^T R B_1 F_0) - B_1^T B_1 R + I \qquad \text{by (C-4)}$$

$$= (B_0 B_1^T R B_1 - B_1^T B_1 R B_0)\, F_0 + I$$

$$= I \qquad \text{by (C-8)} \qquad \text{Q.E.D.}$$

APPENDIX D

In this appendix we want to show that

$$\text{(D-1)} \qquad \lim_{N\to\infty} \left[F_L(N)^* F_L(N)\right]^{-1} = 0$$

when the roots of the characteristic Equation $B(z) = 0$ are on the unit circle. Suppose the roots are distinct, then each f_n can be rewritten as

$$\text{(D-2)} \qquad f_n = \sum_{i=1}^{K} \alpha_i z_i^n = \sum_{i=1}^{K} \alpha_i \exp(j\theta_i n)$$

where $0 \leq \theta_i < 2\pi$

It is easily shown that

$$\text{(D-3)} \qquad \lim \frac{1}{N} \sum_{n=0}^{N} \bar{z}_k^n z_m^n = \lim \frac{1}{N} \sum_{n=0}^{N} \exp[j(\theta_m - \theta_k)n] = \begin{cases} 1 \text{ if } m = k \\ 0 \text{ if } m \neq k \end{cases}$$

Using (D-3), it is easy to show that

$$\text{(D-4)} \qquad \lim \frac{1}{N} F_L^* F_L = \hat{R} = \begin{bmatrix} \hat{r}_0 & \hat{r}_1 & \hat{r}_2 & \cdots & \hat{r}_{k-1} \\ & \hat{r}_0 & \hat{r}_1 & \cdots & \hat{r}_{k-2} \\ & & \ddots & & \vdots \\ & & & \ddots & \vdots \\ \text{hermi-} & & & \ddots & \vdots \\ \text{tian} & & & & \hat{r}_0 \end{bmatrix}$$

where $\hat{r}_m = \sum_{i=1}^{K} \exp(-jm\theta_i)|\alpha_i|^2$, for $m = 0, 1, \ldots, K-1$.

$$\text{Let } Q = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ \exp[j\theta_1] & \exp[j\theta_2] & \cdots & \exp[j\theta_K] \\ \exp[j2\theta_1] & \exp[j2\theta_2] & \cdots & \exp[j2\theta_K] \\ \vdots & \vdots & & \vdots \\ \exp[j(K-1)\theta_1] & \exp[j(K-1)\theta_2] & \cdots & \exp[j(K-1)\theta_K] \end{bmatrix}$$

and

$$A = \text{diag}\,[|\alpha_1|^2, |\alpha_2|^2, \ldots, |\alpha_K|^2], \text{ then}$$

(D-4) is rewritten as

$$\hat{R} = QAQ^* .$$

Because Q is nonsingular and A is positive definite and diagonal, it is shown easily that $\hat{R}$ is positive definite and invertible.

$$\text{Let } \hat{R}(N) = \frac{1}{N} F_L(N)^* F_L(N) .$$

As mentioned in Chapter 3, $F_L(N)^* F_L(N)$ is nonsingular for each N and so is $\hat{R}(N)$.

$$F_L(N)^* F_L(N) = N\hat{R}(N)$$

$$[F_L(N)^* F_L(N)]^{-1} = \frac{1}{N} \hat{R}(N)^{-1}$$

It is shown [7:Ch.4] that if A is nonsingular and $\lim A(k) = A$ then for all sufficiently large k, A(k) is nonsingular and $\lim A(k)^{-1} = A^{-1}$. Because $\lim \hat{R}(N) = \hat{R}$ and $\hat{R}$ is nonsingular,

$$\lim_{N\to\infty} [F_L(N)^* F_L(N)]^{-1} = \lim_{N\to\infty} \frac{1}{N} \hat{R}^{-1} = 0 .$$

REFERENCES

[1] J. L. Melsa and D. L. Cohn, Decision and Estimation Theory, McGraw-Hill, 1978.

[2] C. E. Baum, "Emerging technology for transient and broad-band analysis and synthesis of antennas and scatterers," Proc. IEEE, vol. 64, pp. 1598-1616, Nov. 1976.

[3] F. M. Tesche, "On the analysis of scattering and antenna problems using the singularity expansion technique," IEEE Trans. Antennas and Propagation, vol. AP-21, pp. 53-62, Jan. 1973.

[4] C. E. Baum, "The singularity expansion method: background and development," Electromagnetics, vol. 1, pp. 315-360, Oct.-Dec. 1981.

[5] K. Steiglitz and L. E. McBride, "A technique for the identification of linear systems," IEEE Trans. Automat. Contr., vol. AC-10, pp. 461-464, Oct. 1965.

[6] K. Steiglitz, "On the simultaneous estimation of poles and zeros in speech analysis," IEEE Trans. Acoust., Speech, Signal Processing, vol. ASSP-25, pp. 229-234, June 1977.

[7] S. M. Kay, "Accurate frequency estimation at low signal-to-noise ratio," IEEE Trans. Acoust., Speech, Signal Processing, vol. ASSP-32, pp. 540-547, June 1984.

[8] J. Makhoul, "Linear prediction: A tutorial review," Proc. IEEE, vol. 63, pp. 561-580, Apr. 1975.

[9] K. T. Astrom and P. Eykhoff, "System identification - A survey," Automatica, vol. 7, pp. 123-162, 1971.

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

[11] G. W. Stewart, Introduction of Matrix Computation, New York, Academic Press, 1973.

[12] M. L. Van Blaricum and R. Mittra, "Problems and solutions associated with Prony's method for preprocessing transient data," IEEE Trans. Antennas and Propagation, vol. AP-26, pp. 174-182, Jan. 1978.

[13] B. Noble and J. W. Daniel, Applied Linear Algebra, 2nd edition, Ch.8, N.J., Prentice-Hall, 1977.

[14] R. Kumaresan and D. W. Tufts, "Estimating the parameters of exponentially damped sinusoids and pole-zero modeling in noise," IEEE Trans. Acoust., Speech, Signal Processing, vol. ASSP-30, pp. 883-840, Dec. 1982.

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

[16] T. L. Henderson, "Geometric methods for determining system poles from transient response," IEEE Trans. Acoust., Speech, Signal Processing, vol. ASSP-29, pp. 982-988, Oct. 1981.

[17] P. Stoica and T. Soderstrom, "The Steiglitz - McBride identification algorithm revisited - Convergence analysis and accuracy aspects," IEEE Trans. Automat. Contr., vol. AC-26, pp. 712-717, June 1981.

[18] A. G. Evans and R. Fischl, "Optimal least squares time domain synthesis of recursive digital filters," IEEE Trans. Audio Electroacoust., vol. AU-21, pp. 61-65, Feb. 1973.

[19] C. E. Baum, "The Singularity Expansion Method," Transient Electromagnetic Fields, edited by L. B. Felson, Springer-Verlag, 1976.

[20] K. S. Cho and J. T. Cordaro, "Calculation of the SEM parameters from the transient response of a thin wire," IEEE Trans. Antennas and Propagation, vol. AP-28, pp. 921-924, Nov. 1980.