Statistical Analysis of TLS-Based Prony Techniques

William M. Steedly[†], Chinghui J. Ying [‡] and Randolph L. Moses[‡§]

Suggested running title—Statistical Analysis of TLS-Based Prony.

Subtitle—*We present an analysis of the statistics for the TLS-Prony method by deriving the covariance matrix of the estimated parameters. We verify the theoretical results using Monte-Carlo simulations and compare them to their Cramér-Rao bounds.*

Key Words—Perturbation techniques; Statistics; Parameter estimation; Least-squares estimation.

Abstact

We present an analysis of parameter variance statistics for the TLS-Prony method applied to damped exponential signals. We derive the covariance matrix of the estimated parameters for this method. The parameters include the magnitudes and angles of the poles, and the magnitudes and angles of the amplitude coefficients. We verify the theoretical results using Monte-Carlo simulations studies. We also compare the variance results to the corresponding Cramér-Rao bounds for several cases.

Automatica

Special Issue on Statistical Signal Processing and Control

January 1994

[†]The Analytic Sciences Corporation, Reston, VA 22090.

[‡]Department of Electrical Engineering, The Ohio State University, Columbus, OH 43210.

[§]Author to whom all correspondence should be addressed.

I. Introduction

The problem of estimating model parameters of noisy exponential signals is an active area of research. This problem has applications in a number of areas, including speech processing, deconvolution, radar and sonar signal processing, array processing, and spectrum estimation. A number of authors have considered various aspects of this problem (Schmidt, 1979; Kumaresan and Tufts, 1982b; Roy *et al.*, 1986; Bresler and Macovski, 1986; Rahman and Yu, 1987; Stoica and Nehorai, 1989; Hua and Sarkar, 1990a; Abatzoglou and Lam, 1991, just to name a few), and a large number of algorithms have been developed and analyzed (Henderson, 1981; Kaveh and Barabell, 1986; Rao, 1988; Hua and Sarkar, 1988; Stoica and Nehorai, 1989; Stoica *et al.*, 1989; Ottersten *et al.*, 1991; Stoica and Nehorai, 1991).

One popular class of algorithms for estimating parameters from noisy exponential sequences are the subspace-based approaches. These include the MUSIC algorithm and its enhancements (Schmidt, 1979; Stoica and Nehorai, 1989; Stoica and Nehorai, 1991), subspace rotation methods such as ESPRIT (Roy *et al.*, 1986; Roy and Kailath, 1989; Zoltowski and Stavrinides, 1989; Ottersten *et al.*, 1991; Stoica and Nehorai, 1991), iterative maximum likelihood methods (Bresler and Macovski, 1986; Kumaresan *et al.*, 1986; Ziskind and Wax, 1988), minimum norm methods (Kaveh and Barabell, 1986; Dowling and DeGroat, 1991), and total least squares (TLS) methods (Golub and Van Loan, 1980; Kumaresan and Tufts, 1982a; Rahman and Yu, 1987; Abatzoglou and Lam, 1991). These methods have proven attractive because they exhibit good statistical performance at a modest computational cost. This has been well-established by a large number of numerical studies.

More recently, there has been interest in quantitatively evaluating these methods. To this end, a number of researchers have analyzed the statistical properties of such algorithms (Henderson, 1981; Kaveh and Barabell, 1986; Rao, 1988; Hua and Sarkar, 1988; Stoica and Nehorai, 1989; Stoica *et al.*, 1989; Hua and Sarkar, 1990a; Ottersten *et al.*, 1991; Stoica and Nehorai, 1991). Henderson (1981) presents a geometric study of the pole estimation problem, and analyzes the statistical properties of the prediction coefficients when the Hankel data matrix is corrupted by an i.i.d. noise matrix. Several authors have presented

results relating to pole angle (frequency) estimates from arrays when the exponential signals are undamped, *e.g.*, (Kaveh and Barabell, 1986; Rao, 1988; Stoica and Nehorai, 1989; Stoica *et al.*, 1989; Ottersten *et al.*, 1991; Stoica and Nehorai, 1991; Clergeot *et al.*, 1989). A related perturbation analysis of SVD-based methods is presented in (Vaccaro *et al.*, 1988; Kot *et al.*, 1987; Li and Vaccaro, 1990; Tufts *et al.*, 1991) and applied to both frequency estimation and threshold analysis for exponential modes. Hua and Sarkar (1988) present an angle-only analysis for the least squares Prony method for the poles of undamped exponentials. Less has appeared which considers the statistical properties of the parameters for damped exponentials. Porat and Friedlander consider the related problem of ARMA system identification using SVD-based approaches in (Porat and Friedlander, 1987). Hua and Sarkar (1990a) present an analysis for the pole estimates of damped exponential signals using their matrix pencil method, but have not presented the statistical properties of the amplitude coefficients.

This paper presents an extension of the above works to treat a general exponential case. We introduce a complete statistical derivation for the poles and amplitude coefficients estimated using a TLS-Prony scheme where signals consist of arbitrary damped exponential terms in noise. We provide complete statistics for the individual pole parameters for an exponential model in which the poles may lie on, inside, or outside the unit circle. In addition, we derive the statistical properties of the amplitude coefficients associated with these exponential modes.

The results of this paper provide a sound basis for performance analyses of the TLS-Prony estimation method. We extend previous works by considering the general damped case, as well as by including amplitude coefficient parameters in addition to pole parameters. These results provide the tools to analyze various situations and evaluate the potential success of applying the TLS-Prony estimation algorithm.

The TLS-Prony estimation procedure is a multi-snapshot extension of the algorithm presented in (Kumaresan and Tufts, 1982a; Rahman and Yu, 1987). The advantage of singular value decomposition (SVD) in noise cleaning of the Toeplitz data matrix is well-known. The multiple snapshot incorporation is a straightforward one in which more than one set

of amplitude coefficients corresponds to the set of poles. The procedure is discussed in Section II.

The statistical derivation for this procedure is based on a first order perturbation analysis; thus the analysis assumes high SNR. We derive the complete covariance matrix of the estimated parameters for this case. The parameters include the magnitudes and angles of the poles, and the magnitudes and angles of the amplitude coefficients.

Using these expressions, several general properties of the parameter covariance matrix are derived for the high SNR case. We show that the angle and magnitude parameters are uncorrelated for both the poles and the amplitude coefficients. We also show that if the relative magnitude of the pole or amplitude coefficient estimate is considered (*i.e.* $\frac{\widehat{\alpha}}{\alpha}$, where α is the true magnitude), then the corresponding angle and relative magnitude variances are equal.

This paper also examines pole estimation accuracy as functions of pole magnitude, data length, and pole separation using the variance expressions. We compare these variance results to the corresponding Cramér-Rao bounds and verify the theoretical results using Monte-Carlo simulations. The effects on poles inside and outside the unit circle using backward or forward linear prediction in the TLS-Prony estimation scheme are also detailed.

An outline of this paper is as follows. Section II presents the data model. Section III presents the statistics of the model parameters and their properties. Section IV presents some examples using the statistical expressions. Finally, Section V concludes the paper.

II. Estimation Procedure Review

A. Data Model

Assume we have N "snapshots" of data vectors $y(t)$, each of length m:

$$y(t) = \left[\begin{array}{cccc} y_0(t) & y_1(t) & \cdots & y_{m-1}(t) \end{array} \right]^T \qquad t = 1, 2, \ldots, N. \tag{1}$$

Each data vector is modeled as a noisy exponential sequence

$$y_q(t) = \sum_{i=1}^{n} x_i(t) p_i^q + e_q(t) \qquad q = 0, 1, \ldots, m-1. \tag{2}$$

There are n distinct exponential modes in the data; the n poles $\{p_i\}_{i=1}^n$ do not vary from snapshot to snapshot, but the amplitudes $x_i(t)$ may vary. Here, it is assumed that $\{e_q(t)\}$ are uncorrelated zero mean complex white Gaussian noise sequences with variance σ. Equation (2) may be compactly written as

$$y(t) = Ax(t) + e(t), \tag{3}$$

where $e(t) = \begin{bmatrix} e_0(t) & e_1(t) & \cdots & e_{m-1}(t) \end{bmatrix}^T$, $x(t) = \begin{bmatrix} x_1(t) & x_2(t) & \cdots & x_n(t) \end{bmatrix}^T$, and A is the $m \times n$ Vandermonde matrix derived from n signal poles

$$A = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ p_1 & p_2 & \cdots & p_n \\ p_1^2 & p_2^2 & \cdots & p_n^2 \\ \vdots & \vdots & & \vdots \\ p_1^{m-1} & p_2^{m-1} & \cdots & p_n^{m-1} \end{bmatrix}. \tag{4}$$

B. Parameter Estimation

The multi-snapshot backward linear prediction equations are given by:

$$\begin{bmatrix} y & : & Y \end{bmatrix} \begin{bmatrix} 1 \\ b \end{bmatrix} \approx 0, \tag{5}$$

where

$$b = \begin{bmatrix} b_1 & b_2 & \cdots & b_L \end{bmatrix}^T \tag{6}$$

and where

$$\left[\begin{array}{c:c:c} y & : & Y \end{array}\right] = \left[\begin{array}{ccccc} y_0(1) & y_1(1) & y_2(1) & \cdots & y_L(1) \\ y_1(1) & y_2(1) & y_3(1) & \cdots & y_{L+1}(1) \\ \vdots & \vdots & \vdots & & \vdots \\ y_{m-(L+1)}(1) & y_{m-L}(1) & y_{m-(L-1)}(1) & \cdots & y_{m-1}(1) \\ \hline y_0(2) & y_1(2) & y_2(2) & \cdots & y_L(2) \\ y_1(2) & y_2(2) & y_3(2) & \cdots & y_{L+1}(2) \\ \vdots & \vdots & \vdots & & \vdots \\ y_{m-(L+1)}(2) & y_{m-L}(2) & y_{m-(L-1)}(2) & \cdots & y_{m-1}(2) \\ \hline & & \vdots & & \\ \hline y_0(N) & y_1(N) & y_2(N) & \cdots & y_L(N) \\ y_1(N) & y_2(N) & y_3(N) & \cdots & y_{L+1}(N) \\ \vdots & \vdots & \vdots & & \vdots \\ y_{m-(L+1)}(N) & y_{m-L}(N) & y_{m-(L-1)}(N) & \cdots & y_{m-1}(N) \end{array}\right]. \tag{7}$$

Here L is the order of prediction and b is the coefficient vector of the polynomial $B(z)$ given by

$$B(z) = 1 + b_1 z^1 + b_2 z^2 + \cdots + b_L z^L. \tag{8}$$

For the noiseless case, L can be any integer greater than or equal to the model order n; however, in the presence of noise choosing $L > n$ results in more accurate parameter estimates (see Section II). Note that all of the N snapshots are used simultaneously to estimate a single set of prediction coefficients (and therefore, a single set of poles).

The TLS-Prony method considers the effect of noise perturbation of both Y and y, and the TLS solution attempts to minimize the effect of these perturbations on the prediction coefficient vector b (see (Kumaresan and Tufts, 1982a; Rahman and Yu, 1987) for details). This is accomplished by obtaining an SVD of the matrix $\left[\begin{array}{c:c} y & Y \end{array}\right]$ and truncating all but the first n singular values to arrive at an estimate $\left[\begin{array}{c:c} \widehat{y} & \widehat{Y} \end{array}\right]$ (Kumaresan and Tufts, 1982a; Rahman and Yu, 1987). Inserting $\left[\begin{array}{c:c} \widehat{y} & \widehat{Y} \end{array}\right]$ in Equation (5) gives the modified linear prediction equation

$$\widehat{Y}\widehat{b} = -\widehat{y} \tag{9}$$

from which the linear prediction coefficient vector estimate $\widehat{b}$ is found as

$$\widehat{b} = -\widehat{Y}^{+}\widehat{y}, \tag{10}$$

where $^{+}$ denotes the Moore-Penrose pseudoinverse. A numerically robust solution for $\widehat{b}$ can be found directly from the SVD of $\left[\begin{array}{c:c} \widehat{y} & \widehat{Y} \end{array}\right]$, as is shown in (Rahman and Yu, 1987). Finally, the estimates for the poles are found by

$$\widehat{p}_j = \mathrm{zero}_j\left(\widehat{B}(z)\right), \qquad j = 1, 2, \ldots, L. \tag{11}$$

Once the L poles are determined from Equation (11), one must separate the n "true" poles from the remaining $L - n$ "extraneous" or "noise" poles. A popular approach is to choose n poles based on their location with respect to the unit circle. For example, one can choose the n poles closest to the unit circle if it is known that the poles are undamped (Stoica and Söderdtröm, 1991) or the n poles with smallest moduli if it is known that the poles are damped (Kumaresan and Tufts, 1982b). However, these methods do not apply when the true poles may lie both inside and outside the unit circle. In this case we can classify poles as true or extraneous based on the energy of the corresponding mode. We have found this method to be more reliable than other procedures for the case when the true modes lie both inside and outside the unit circle. This arises, for example, in the radar scattering problem where measurements are made over a small relative bandwidth, and the exponential modes in the data can be decaying or growing over that band (Sacchini *et al.*, 1992; Sacchini, 1992).

In this energy criterion method, the L sets of amplitude coefficients can be found using the pole estimates, and Equation (3) leads to the following least squares equation for the amplitude coefficients,

$$\begin{bmatrix} 1 & 1 & \cdots & 1 \\ \widehat{p}_1 & \widehat{p}_2 & \cdots & \widehat{p}_L \\ \widehat{p}_1^2 & \widehat{p}_2^2 & \cdots & \widehat{p}_L^2 \\ \vdots & \vdots & & \vdots \\ \widehat{p}_1^{m-1} & \widehat{p}_2^{m-1} & \cdots & \widehat{p}_L^{m-1} \end{bmatrix} \begin{bmatrix} \widehat{x}_1(1) & \widehat{x}_1(2) & \cdots & \widehat{x}_1(N) \\ \widehat{x}_2(1) & \widehat{x}_2(2) & \cdots & \widehat{x}_2(N) \\ \vdots & \vdots & & \vdots \\ \widehat{x}_L(1) & \widehat{x}_L(2) & \cdots & \widehat{x}_L(N) \end{bmatrix} \approx \begin{bmatrix} y(1) & y(2) & \cdots & y(N) \end{bmatrix} \tag{12}$$

or

$$\widehat{A}_L \widehat{X}_L \approx Y_a. \tag{13}$$

A least squares solution to Equation (13) can be computed as

$$\widehat{X}_L = \left(\widehat{A}_L^* \widehat{A}_L\right)^{-1} \widehat{A}_L^* Y_a = \widehat{A}_L^+ Y_a, \tag{14}$$

where * denotes complex conjugate transpose. (However, in practice, more numerically robust procedures, such as a QR decomposition, should be used to solve Equation (13).) Because only n singular values of $\left[\begin{array}{c:c} \widehat{y} & \widehat{Y} \end{array}\right]$ are nonzero, there are at most n pole estimates which can correspond to data modes. Therefore, only the n poles which have the largest energy are retained. This is done by computing the L mode energies as

$$E_i = \sum_{t=1}^{N} |\widehat{x}_i(t)|^2 \sum_{q=0}^{m-1} |\widehat{p}_i|^{2q} \qquad i = 1, 2, \ldots, L \tag{15}$$

and retaining those n poles whose corresponding energies are highest. We then reestimate the amplitude coefficients of these n poles. This is done using Equation (14), except that $\widehat{A}_L$ is replaced by $\widehat{A}$, where $\widehat{A}$ is the Vandermonde matrix composed only of the n columns of $\widehat{A}_L$ corresponding to the n selected poles. We note that the above procedure produces consistent estimates as $\sigma \to 0$, as is shown in the Appendix.

III. Statistical Analysis

In this section we present a first order statistics of the complete set of parameter estimates obtained in the TLS-Prony method. Assume the data is given as in Equation (2). Let ω_i and α_i be the angle and magnitude, respectively, of each pole p_i, thus $p_i = \alpha_i e^{j\omega_i}$. Similarly let $\gamma(t)$ and $\beta(t)$ be the angle and magnitude vectors, respectively, of each vector of amplitude coefficients $x(t)$,

$$\begin{aligned} \gamma(t) &= \left[\begin{array}{cccc} \gamma_1(t) & \gamma_2(t) & \cdots & \gamma_n(t) \end{array}\right]^T \\ \beta(t) &= \left[\begin{array}{cccc} \beta_1(t) & \beta_2(t) & \cdots & \beta_n(t) \end{array}\right]^T, \end{aligned} \tag{16}$$

where

$$x(t) = \left[\begin{array}{cccc} \beta_1(t)e^{j\gamma_1(t)} & \beta_2(t)e^{j\gamma_2(t)} & \cdots & \beta_n(t)e^{j\gamma_n(t)} \end{array} \right]^T . \tag{17}$$

Define following parameter vectors:

$$\begin{aligned} \theta_x &= \left[\begin{array}{cccccccc} \gamma^T(1) & \beta^T(1) & \gamma^T(2) & \beta^T(2) & \cdots & \gamma^T(N) & \beta^T(N) \end{array} \right]^T \\ \theta_p &= \left[\begin{array}{cccccccc} \omega_1 & \omega_2 & \cdots & \omega_n & \alpha_1 & \alpha_2 & \cdots & \alpha_n \end{array} \right]^T \\ \theta &= \left[\begin{array}{cc} \theta_x^T & \theta_p^T \end{array} \right]^T . \end{aligned} \tag{18}$$

The following theorem gives the first order approximation of the probability density function (pdf) of $\widehat{\theta}$.

Theorem 1: Let $\widehat{\theta}$ denote the TLS-Prony estimate of θ which is given by the n highest energy mode estimates found in Equations (11) and (14). Then to a first order approximation (as $\sigma \to 0$), the pdf of $\widehat{\theta}$ is given by the unbiased Normal distribution

$$\widehat{\theta} \sim \mathcal{N}(\theta, \Sigma_\theta), \tag{19}$$

where

$$\Sigma_\theta = \left[\begin{array}{cccccc} \overline{U}(1,1) & \widetilde{U}(1,1)T_\beta{}^{-1}(1) & \cdots & \overline{U}(1,N) & \widetilde{U}(1,N)T_\beta{}^{-1}(N) & \overline{V}(1) & \widetilde{V}(1)T_\alpha{}^{-1} \\ -T_\beta^{-1}(1)\widetilde{U}(1,1) & T_\beta^{-1}(1)\overline{U}(1,1)T_\beta{}^{-1}(1) & \cdots & -T_\beta^{-1}(1)\widetilde{U}(1,N) & T_\beta^{-1}(1)\overline{U}(1,N)T_\beta{}^{-1}(N) & -T_\beta^{-1}(1)\widetilde{V}(1) & T_\beta^{-1}(1)\overline{V}(1)T_\alpha{}^{-1} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ \overline{U}(N,1) & \widetilde{U}(N,1)T_\beta{}^{-1}(1) & \cdots & \overline{U}(N,N) & \widetilde{U}(N,N)T_\beta{}^{-1}(N) & \overline{V}(N) & \widetilde{V}(N)T_\alpha{}^{-1} \\ -T_\beta^{-1}(N)\widetilde{U}(N,1) & T_\beta^{-1}(N)\overline{U}(N,1)T_\beta{}^{-1}(1) & \cdots & -T_\beta^{-1}(N)\widetilde{U}(N,N) & T_\beta^{-1}(N)\overline{U}(N,N)T_\beta{}^{-1}(N) & -T_\beta^{-1}(N)\widetilde{V}(N) & T_\beta^{-1}(N)\overline{V}(N)T_\alpha{}^{-1} \\ \overline{V^*}(1) & \widetilde{V^*}(1)T_\beta{}^{-1}(1) & \cdots & \overline{V^*}(N) & \widetilde{V^*}(N)T_\beta{}^{-1}(N) & \overline{Z} & \widetilde{Z}T_\alpha{}^{-1} \\ -T_\alpha^{-1}\widetilde{V^*}(1) & T_\alpha^{-1}\overline{V^*}(1)T_\beta{}^{-1}(1) & \cdots & -T_\alpha^{-1}\widetilde{V^*}(N) & T_\alpha^{-1}\overline{V^*}(N)T_\beta{}^{-1}(N) & -T_\alpha^{-1}\widetilde{Z^*} & T_\alpha^{-1}\overline{Z}T_\alpha{}^{-1} \end{array} \right], \tag{20}$$

where $\bar{\cdot}$ and $\tilde{\cdot}$ in Equation (20) are real and imaginary part operators, respectively, and where $U(t,r)$, $V(t)$, and Z are $n \times n$ complex matrices which depend on θ, L, and m. The specific formulas for these entries can be found in the Appendix. $T_\beta(t)$ and T_α are diagonal matrices given by

$$T_\beta(t) = \mathrm{diag}\left(\frac{1}{\beta_1(t)}, \frac{1}{\beta_2(t)}, \ldots, \frac{1}{\beta_n(t)} \right)$$

$$T_\alpha \;=\; \mathrm{diag}\left(\frac{1}{\alpha_1}, \frac{1}{\alpha_2}, \ldots, \frac{1}{\alpha_n}\right). \tag{21}$$

Proof: See the Appendix. □

Several properties of the covariance can be derived from the structure of Σ_θ. Some of these properties are presented in the following corollaries.

Corollary 1: From Σ_θ in Equation (20), $\mathrm{Cov}\left(\widehat{\gamma}_i(t), \widehat{\beta}_i(t)\right) = 0$, and $\mathrm{Cov}\left(\widehat{\omega_i}, \widehat{\alpha_i}\right) = 0$.

Proof: Consider the blocks of Σ_θ containing the covariances of interest, which are given by $\widetilde{U}(t,t){T_\beta}^{-1}(t)$ and $\widetilde{Z}{T_\alpha}^{-1}$. From Equation (66) in the Appendix, it can be seen that $U(t,t)$ and Z are Hermitian. If follows that the diagonal elements of $\widetilde{U}(t,t)$ and $\widetilde{Z}$ are zero. Since $T_\beta(t)$ and T_α are real, diagonal matrices, the diagonal elements of $\widetilde{U}(t,t){T_\beta}^{-1}(t)$ and $\widetilde{Z}{T_\alpha}^{-1}$ are also zero, which gives the desired result. □

Note that when $t \neq r$, $U(t,r)$ is not Hermitian, so the diagonal elements of $\widetilde{U}(t,r)$ are not zero. Thus it is not in general the case that the magnitude of $x_i(t)$ and the angle of $x_j(t)$ are uncorrelated for $i \neq j$.

Note that from Equation (20) the angle variances are equal to the magnitude variances except for the transformation matrices $T_\beta(t)$ and T_α. These transformation matrices can be eliminated by rescaling some of the parameters in θ. The required rescaling is obtained by using the *relative* magnitudes of the poles and amplitude coefficients as the estimated parameters instead of their absolute magnitudes. That is, define the estimate $\widehat{\theta}_1$ to be as in Equation (18), but with $\widehat{\alpha}_i$ and $\widehat{\beta}_i(t)$ replaced by the relative magnitudes $\frac{\widehat{\alpha}_i}{\alpha_i}$ and $\frac{\widehat{\beta}_i(t)}{\beta_i(t)}$. We then note that the Jacobian transformation from θ to θ_1 is given by

$$J = \mathrm{diag}\left(I_n, T_\beta(1), I_n, T_\beta(2), \ldots, I_n, T_\beta(N), I_n, T_\alpha\right). \tag{22}$$

Corollary 2: $\Sigma_{\theta_1} = \mathrm{cov}\left(\widehat{\theta}_1\right)$ is given by Equation (20) with all $T_\beta(t)$ and T_α matrices replaced by identity matrices. It follows that the covariances of parameter angles are equal to the covariances of the corresponding relative magnitudes.

Proof: Immediate from the fact that $\Sigma_{\theta_1} = J\Sigma_\theta J$ with J defined in Equation (22). □

We can also consider a reparameterization of θ in which real and imaginary parts of the amplitude coefficients and poles are considered as parameters. Let us denote such a reparameterization as θ_2, with corresponding covariance matrix which would give Σ_{θ_2}.

Corollary 3: Let ν denote a complex parameter, which is either a pole p_i or an amplitude coefficient $x_i(t)$. Then $\text{var}(\text{Re}\{\nu\}) = \text{var}(\text{Im}\{\nu\})$ and $\text{Re}\{\nu\}$ is uncorrelated with $\text{Im}\{\nu\}$.

Proof: The result can be obtained by applying the Jacobian variable transformation from polar to rectangular coordinates to Σ_θ. This transformation is straightforward, but tedious, and not presented here. □

Corollary 4: Σ_θ is independent of the absolute phase references of the amplitude coefficients within each snapshot, $\phi(t)$, and independent of the absolute phase reference of the poles, ϕ.

Proof: The result follows by examining the expressions for $T_\beta(t)$ and T_α in Equation (21), $U(t,r)$, $V(t)$, and Z in the Appendix, and noting that they remain unchanged if $T_x(t)$ is replaced by $e^{j\phi(t)}T_x(t)$ and p_i is replaced by $e^{j\phi}p_i$. □

IV. Examples

In this section we present a set of examples which illustrate the performance of the TLS-Prony method. We first compare the first order statistics presented above to the CRB for a number of cases. The CRB for this data model is presented in (Steedly and Moses, 1991). We then compare the first order statistics to those obtained using Monte-Carlo simulation.

A. Example 1: Single Exponential Mode

In this example we consider a single pole model with one snapshot of data (and thus one amplitude coefficient). The experiment entails moving the pole along the positive real axis from 0.1 to 10, *i.e.* $0.1 \leq p \leq 10$ (the results are independent of the pole angle by Corollary 4, so an angle of zero is chosen). For each pole location , we calculate the parameter variances using Equation (20) for data sets of lengths 2, 5, 10, 20, 50, and 100. For comparative purposes, the amplitude coefficient associated with the pole is chosen to be a positive real number such that the mode energy ($x^2 \sum_{l=0}^{m-1} p^{2l}$) is unity for each pole location and data length. The noise power is also kept constant at $\sigma = 1$. The model order L is chosen to be one third of the data length m, which has been shown to be near optimal for a number of cases (see (Hua and Sarkar, 1990a; Steedly *et al.*, 1992), and *A*.3 below).

1. Pole Variances

The first order theoretical variances for the estimated pole angle and pole magnitude appear as the dashed lines in Figures 1 and 2, respectively; the corresponding CRBs appear as the solid lines in these figures. From Figure 1 we see that the pole angle variances are close to the CRBs when the pole is inside the unit circle. When the pole is outside the unit circle, the variances become much higher than the CRBs (except for the $m = 2$ case). For larger data lengths the disparity with the CRB is much more pronounced. This is because backward linear prediction is used in our TLS-Prony estimation method. With backward linear prediction, extraneous poles lie outside the unit circle, thus making estimation of poles outside the unit circle more difficult (Kumaresan, 1983). The use of forward linear prediction would give corresponding results for poles inside the unit circle. Similar observations apply to the pole magnitude variances (see Figure 2). The pole magnitude variances can be normalized to give relative error of the pole magnitude, *i.e.* $\mathrm{var}\left(\frac{\widehat{\alpha}_1}{\alpha_1}\right)$. If this is done, one obtains exactly the same curves as in Figure 1 (*cf.* Corollary 2).

From these two figures we also see that inside the unit circle the variances for pole angle are higher than the variances for pole magnitude and vice-versa outside the unit circle. This is because the angular uncertainty becomes greater as a pole moves closer to the origin. As expected, the pole angle variance approaches infinity as the pole approaches the origin.

From Figures 1 and 2 we see that the pole angle and magnitude variances are asymptotically (as $m \to \infty$) lowest when the pole is on the unit circle, and that on the unit circle the variances are decreasing by $1/m^2$ (m is the data length). This is consistent with the well-known $1/m^3$ variance decrease, since the amplitude coefficient is adjusted in this experiment to keep the mode energy constant (if the amplitude coefficient is left unchanged, the variance decrease is $1/m^3$).

When the pole is not on the unit circle, the variances do not decrease to zero as $m \to \infty$. Recall that we keep the total mode energy constant. For a decaying or growing exponential mode, adding data points while keeping the energy constant results in adding data points with smaller and smaller amplitude. As a result, the parameter estimate variances do not continue to decrease.

2. Amplitude Coefficient Variances

The variances for the amplitude coefficient angle and magnitude appear in Figures 3 and 4. As before, each curve is a plot of variance versus pole radius for various data lengths m. There are several points to note in these figures. First, when the pole is inside the unit circle, increasing the number of data points provides no significant decrease in the variances. The first data point with no noise added is precisely the amplitude coefficient. When the pole is inside the unit circle, this amplitude does not change much as a function of data length, and consequently its variance does not change much. However, when the pole is outside the unit circle, the amplitude coefficient decreases rapidly toward zero as data length increases. Thus, outside the unit circle the estimate of the amplitude cannot be expected to vary much around zero and the magnitude variances become low. Also, variance of the estimated angle becomes quite large because of high angular uncertainty for points near zero.

3. Prediction Order Considerations

In the above figures we used a prediction order L equal to one-third the data length. We next consider the effect of prediction order on the variances of the TLS-Prony parameter estimates. We consider a single exponential mode whose pole is on the unit circle. We choose $\sigma = 1$ and choose the amplitude coefficient so that the mode energy is unity, as before.

Figures 5, 6, and 7 show the variances of the pole angle, amplitude coefficient magnitude, and amplitude coefficient angle. The solid lines represent the CRBs (which are, of course, independent of TLS prediction order), and the dotted lines represent the TLS-Prony variances. Figure 5 has been presented in earlier works (Hua and Sarkar, 1990a; Clergeot *et al.*, 1989; Kot *et al.*, 1987), but the amplitude coefficient was not considered there. Since the pole is on the unit circle, the pole magnitude results are identical to the pole angle results (*cf.* Corollary 2, with $\alpha_1 = 1$). From these figures we can see that the best prediction order choice is approximately $m/3$; this agrees with the results in (Hua and Sarkar, 1990a; Clergeot *et al.*, 1989; Kot *et al.*, 1987).

B. *Example 2: Two Undamped Exponential Modes*

In this example we consider two poles at $\alpha_1 e^{j(\omega_0+\Delta\omega/2)}$ and $\alpha_2 e^{j(\omega_0-\Delta\omega/2)}$ with $\alpha_1 = \alpha_2 = 1$. Variances are computed for various data lengths ($m = 5$, 10, 20, 50, and 100) and angle separations $\Delta\omega$. The variance results are independent of ω_0 so $\omega_0 = 0$ is used. Again, $L = m/3$, $\sigma = 1$, one snapshot of data is used ($N = 1$), and each amplitude is chosen such that the corresponding mode energy is unity.

Figure 8 shows the variances for the pole angle estimate (of either pole) as a function of pole separation and data length. The solid lines again show corresponding CRB results. The variances for the pole magnitudes are equal to the pole angle variances because these poles are located on the unit circle. We see that the TLS-Prony algorithm performs well with respect to the CRB. In fact, for the curves shown, the TLS-Prony variance curves are always within 2dB of the corresponding CRB curves.

C. *Example 3: Monte-Carlo Simulation of Ten Mode Case*

In this example, we have chosen ten exponential modes to represent a general case. The true pole location of each mode is denoted with an "x" in Figure 9(a). For this case we have $m = 100$ data points, $L = 33$, and $\sigma = 0.001$. The amplitude coefficients are chosen so that each mode energy is unity; this corresponds to an SNR of 10dB per mode; the total SNR (signal power/noise power) is 20dB. The modes were chosen as a representative sampling of various situations.

Figure 9 presents a comparison between the TLS-Prony estimate theoretical variances and variance estimates obtained using Monte-Carlo simulations. The theoretical variances are shown as two-standard deviation concentration ellipses around each pole. These ellipses (they are actually circles, by Corollary 3) give 87% confidence intervals in the complex plane for pole pole estimates. The dots in Figure 9 are pole estimates corresponding to the ten highest energy poles from each of 100 Monte-Carlo simulations. The details of the pole estimation accuracy are summarized in Table 1. Note that the Monte-Carlo variances are close to those predicted by theory for most of the poles, in particular for those closer to the unit circle. For poles well inside the unit circle, there is some bias present which is not

predicted by a first order approximation; in addition, the predicted variance is smaller than the actual variance. As poles move outside the unit circle to the radius of the extraneous poles, some deterioration occurs in terms of misidentifying pole estimates as "true" or "extraneous". Note the rapid increase in variance of a pole estimate as its radius increases, by comparing the variance for pole 2, 8, 4, and 1.

V. Conclusions

We have presented a statistical analysis for estimated poles of the TLS-Prony algorithm. This analysis includes complete expressions for the covariance matrix of the parameters of an exponential model which contains one set of poles and multiple sets of amplitude coefficients. The poles of this model may lie anywhere in the complex plane. Using these expressions, several useful properties of the covariance matrix were established. These include independence of the two parameters for each amplitude coefficient and pole, whether one considers a polar, a relative magnitude polar, or a rectangular real and imaginary part parameterization. It was also established that the variances of these pairs are equal for the relative magnitude polar and rectangular real and imaginary part parameterizations.

The results of this paper provide a sound basis for performance analyses of the TLS-Prony estimation method. We have extended previous works to include the general damped undamped case, as well as to include amplitude coefficient parameters in addition to pole parameters. The results can be used to analyze various situations and evaluate the potential success of applying the TLS-Prony estimation algorithm, as the corollaries and examples in the paper demonstrate.

Acknowledgements

This research was supported by the Air Force Office of Scientific Research, the Avionics Division, Wright Laboratories, and the Surveillance Division, Rome Laboratories.

References

[Abatzoglou and Lam, 1991] T. J. Abatzoglou and L. K. Lam. Direction finding using uniform arrays and the constrained total least squares method. In *Proceedings of the Twenty-Fifth Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, November 4–6 1991.

[Bresler and Macovski, 1986] Y. Bresler and A. Macovski. Exact maximum likelihood parameter estimation of superimposed exponential signals in noise. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, ASSP–34(5):1361–1375, October 1986.

[Clergeot *et al.*, 1989] H. Clergeot, S. Tressens, and A. Ouamri. Performance of high resolution frequencies estimation methods compared to the Cramér-rao bounds. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, ASSP-37(11):1703–1720, November 1989.

[Dowling and DeGroat, 1991] E. M. Dowling and R. D. DeGroat. The equivalence of the total least squares and minimum norm methods. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, ASSP-39(8):1891–1892, August 1991.

[Golub and Van Loan, 1980] G. H. Golub and C. F. Van Loan. An analysis of the total least squares problem. *SIAM*, 17(6):883–893, December 1980.

[Henderson, 1981] T. L. Henderson. Geometric methods for determining system poles from transient response. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, ASSP-29(5):982–988, October 1981.

[Hua and Sarkar, 1988] Y. Hua and T. K. Sarkar. Perturbation analysis of TK method for harmonic retrieval problems. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, ASSP-36(2):228–240, February 1988.

[Hua and Sarkar, 1990a] Y. Hua and T. K. Sarkar. Matrix pencil method for estimating parameters of exponentially damped/undamped sinusoids in noise. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, ASSP-38(5):814–824, May 1990.

[Hua and Sarkar, 1990b] Y. Hua and T. K. Sarkar. A perturbation property of the TLS-LP method. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, ASSP-38(11):2004–2005, November 1990.

[Kaveh and Barabell, 1986] M. Kaveh and A. J. Barabell. The statistical performance of the MUSIC and the minimum-norm algorithms in resolving plane waves in noise. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, ASSP–34(2):331–341, April 1986.

[Kot *et al.*, 1987] A. Kot, S. Parthasarathy, D. Tufts, and R. Vaccaro. Statistical performance of single sinusoid frequency estimation in white noise using state-variable balancing and linear prediction. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, ASSP-35(11):1639–1642, November 1987.

[Kumaresan and Tufts, 1982a] R. Kumaresan and D. Tufts. Accurate parameter estimation of noisy speech-like signals. In *Proc. Internat. Conf. Acoust. Speech. Signal Process. '82*, volume 3, pages 1357–1361, 1982.

[Kumaresan and Tufts, 1982b] R. Kumaresan and D. W. Tufts. Estimating the parameters of exponentially damped sinusoids and pole-zero modeling in noise. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, ASSP-30(6):833–840, December 1982.

[Kumaresan *et al.*, 1986] R. Kumaresan, L. L. Scharf, and A. K. Shaw. An algortihm for pole-zero modeling and analysis. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, ASSP–34(6):637–640, June 1986.

[Kumaresan, 1983] R. Kumaresan. On the zeros of the linear prediction-error filter for deterministic signals. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, ASSP-31(1):217–220, February 1983.

[Li and Vaccaro, 1990] F. Li and R. Vaccaro. Analysis of min-norm and MUSIC with arbitrary array geometry. *IEEE Transactions on Aerospace and Electronic Systems*, AES-38(6):976–985, November 1990.

[Ottersten *et al.*, 1991] B. Ottersten, M. Viberg, and T. Kailath. Performance analysis of the total least squares ESPRIT algorithm. *IEEE Transactions on Signal Processing*, SP-39(5):1122–1135, May 1991.

[Porat and Friedlander, 1987] B. Porat and B. Friedlander. On the accuracy of the Kumaresan-Tufts method for estimating complex damped exponentials. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, ASSP-35(2):231–235, February 1987.

[Rahman and Yu, 1987] M. A. Rahman and K.-B. Yu. Total least squares approach for frequency estimation using linear prediction. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, ASSP–35(10):1440–1454, October 1987.

[Rao, 1988] B. .D. Rao. Perturbation analysis of an SVD-based linear prediction method for estimating the frequencies of multiple sinusoids. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, ASSP-36(7):1026–1035, July 1988.

[Roy and Kailath, 1989] R. Roy and T. Kailath. ESPRIT– estimation of signal parameters via rotational invariance techniques. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, ASSP-37(7):984–995, July 1989.

[Roy *et al.*, 1986] R. Roy, A. Paulraj, and T. Kailath. ESPRIT–a subspace rotation approach to estimation of parameters of cisoids in noise. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, ASSP-34(4):1340–1342, April 1986.

[Sacchini *et al.*, 1992] J. Sacchini, W. M. Steedly, and R. L. Moses. Two-dimensional Prony modeling and parameter estimation. In *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, pages III: 333–336, San Francisco, April 1992.

[Sacchini, 1992] J. J. Sacchini. *Development of Two-Dimensional Parametric Radar Signal Modeling and Estimation Techniques with Application to Target Identification.* PhD thesis, The Ohio State University, March 1992.

[Schmidt, 1979] R. O. Schmidt. Multiple emitter location and signal parameter estimation. In *Proc. RADC Spectral Estimation Workshop*, pages 243–258, Rome, NY, 1979.

[Steedly and Moses, 1991] W. M. Steedly and R. L. Moses. The Cramér-Rao bound for pole and amplitude estimates of damped exponential signals in noise. In *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, pages 3569–3572, Toronto, Ontario, May 14–17, 1991.

[Steedly *et al.*, 1992] W. M. Steedly, C. J. Ying, and R. L. Moses. A modified TLS-Prony method using data decimation. In *Proceedings of the SPIE OE/AEROSPACE Science and Sensing Conference*, Orlando, FL, April 22–24, 1992.

[Stoica and Nehorai, 1989] Petre Stoica and Arye Nehorai. MUSIC, Maximum Likelihood, and Cramér-Rao Bound. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, ASSP–37(5):720–741, May 1989.

[Stoica and Nehorai, 1991] P. Stoica and A. Nehorai. Performance comparison of subspace rotation and MUSIC methods for direction estimation. *IEEE Transactions on Signal Processing*, SP-31(2):446–453, February 1991.

[Stoica and Söderdtröm, 1991] P. Stoica and T. Söderdtröm. On spectral and root forms of sinusoidal frequency estimators. In *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, pages 3257–3260, Toronto, Ontario, May 14–17, 1991.

[Stoica *et al.*, 1989] P. Stoica, T. Söderström, and F.-N. Ti. Asymptotic properties of the high-order Yule-Walker estimates of sinusoidal frequencies. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, ASSP–37(11):1721–1734, November 1989.

[Tufts *et al.*, 1991] D. Tufts, A. Kot, and R. Vaccaro. The threshold effect in signal processing algorithms which use an estimated subspace. In *SVD and Signal Processing, II–Algorithms, Analysis and Applications (edited by R. Vaccaro)*, pages 301–320. Elsevier Science Publishers B.V.(North-Holland), 1991.

[Vaccaro *et al.*, 1988] R. Vaccaro, D. Tufts, and G. Boudreaux-Bartels. Advances in principal component signal processing. In *SVD and Signal Processing — Algorithms, Applications and architectures (edited by E. Deprettere)*, pages 115–146. Elsevier Science Publishers B.V.(North-Holland), 1988.

[Wedin, 1973] P.-A. Wedin. Perturbation theory for pseudo-inverses. *BIT*, 13:217–232, 1973.

[Ziskind and Wax, 1988] I. Ziskind and M. Wax. Maximum likelihood localization of multiple sources by alternating projection. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, ASSP–36(10):1553–1560, October 1988.

[Zoltowski and Stavrinides, 1989] M. D. Zoltowski and D. Stavrinides. Sensor array signal processing via a Procrustes rotations based eigenanalysis of the ESPRIT data pencil. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, ASSP-37(6):832–861, June 1989.

Appendix: Proof of Theorem 1

From Equation (10) we note that

$$\begin{aligned}\widetilde{b} \stackrel{\triangle}{=} \widehat{b} - b &= -\widehat{Y}^{+}\widehat{y} + S^{+}s \\ &= -\widehat{Y}^{+}(s+\widetilde{s}) + S^{+}s \\ &= -\left(\widehat{Y}^{+} - S^{+}\right)s - \widehat{Y}^{+}\widetilde{s} \end{aligned} \tag{23}$$

where

$$\left[\begin{array}{ccc} \widehat{y} & : & \widehat{Y} \end{array}\right] = \left[\begin{array}{ccc} s & : & S \end{array}\right] + \left[\begin{array}{ccc} \widetilde{s} & : & \widetilde{S} \end{array}\right]. \tag{24}$$

Here, $\left[\begin{array}{ccc} s & : & S \end{array}\right]$ is the noise free version of $\left[\begin{array}{ccc} \widehat{y} & : & \widehat{Y} \end{array}\right]$, and b is the Lth order linear prediction coefficient vector associated with the poles of the noiseless model. Note that b gives the n true poles and $L-n$ extraneous poles (*cf.* (Kumaresan, 1983)). In order to evaluate the expression in Equation (23) we use the following identity for any matrix M (Wedin, 1973)

$$\widehat{M}^{+} - M^{+} = -\widehat{M}^{+}\widetilde{M}M^{+} + \left(\widehat{M}^{*}\widehat{M}\right)^{+}\widetilde{M}^{*}P_{M}^{\perp} + P_{\widehat{M}^{*}}^{\perp}\widetilde{M}^{*}\left(MM^{*}\right)^{+}, \tag{25}$$

where $\widetilde{M} = \widehat{M} - M$, $P_{M}^{\perp} = I_m - MM^{+}$, $P_{\widehat{M}^{*}}^{\perp} = I_n - \widehat{M}^{*}\widehat{M}^{*^{+}}$, and I_q is the $q \times q$ identity matrix. Using the fact that $P_{S}^{\perp}s = 0$ and a first order approximation, we then obtain

$$\begin{aligned}\widetilde{b} &= -\widehat{Y}^{+}\widetilde{S}b - P_{\widehat{Y}^{*}}^{\perp}\widetilde{S}^{*}S^{*^{+}}S^{+}s - \widehat{Y}^{+}\widetilde{s} \\ &\approx -S^{+}\widetilde{S}b - P_{S^{*}}^{\perp}\widetilde{S}^{*}S^{*^{+}}S^{+}s - S^{+}\widetilde{s}. \end{aligned} \tag{26}$$

The above approximation is valid since the matrices $\left[\begin{array}{ccc} \widehat{y} & : & \widehat{Y} \end{array}\right]$ and $\left[\begin{array}{ccc} s & : & S \end{array}\right]$ have the same rank. Let

$$\left[\begin{array}{ccc} y & : & Y \end{array}\right] = \left[\begin{array}{ccc} s & : & S \end{array}\right] + \left[\begin{array}{ccc} w & : & W \end{array}\right], \tag{27}$$

where $\left[\begin{array}{ccc} w & : & W \end{array}\right]$ is defined as the noise-only part of $\left[\begin{array}{ccc} y & : & Y \end{array}\right]$ (see Equations (2) and (7)). By using the perturbation analysis in (Hua and Sarkar, 1990b) on the matrices

$\left[\begin{array}{ccc} s & : & S \end{array}\right]$, $\left[\begin{array}{ccc} y & : & Y \end{array}\right]$, and $\left[\begin{array}{ccc} \widehat{y} & : & \widehat{Y} \end{array}\right]$, it can be shown that to a first order approximation

$$S^{+}\left[\begin{array}{ccc} \widetilde{s} & : & \widetilde{S} \end{array}\right] = S^{+}\left[\begin{array}{ccc} w & : & W \end{array}\right]. \tag{28}$$

The above equation also implies that $\widetilde{S}^{*}S^{*+} = W^{*}S^{*+}$. Thus, Equation (26) can be written (to a first order approximation) as

$$\widetilde{b} = -S^{+}Wb - P_{\widetilde{S}^{*}}^{\perp}W^{*}S^{*+}S^{+}s - S^{+}w. \tag{29}$$

From Equation (29), we note that $\widetilde{b} \to 0$ as $\sigma \to 0$ since the elements of W and w are uncorrelated, zero mean, complex white Gaussian random variables. Therefore, the resulting L pole estimates are consistent as $\sigma \to 0$. Similarly it can be shown that the L sets of the amplitude coefficients are also consistent as $\sigma \to 0$. Note that the "true" amplitude coefficients of the extraneous modes are zero; thus it follows that choosing the n highest energy poles as the true poles is consistent as $\sigma \to 0$.

Note that $SP_{\widetilde{S}^{*}}^{\perp} = 0$. Multiplying both sides of Equation (29) through by S, we obtain

$$S\widetilde{b} = -SS^{+}\epsilon, \tag{30}$$

where $\epsilon = w + Wb$.

From Equation (2), we can write S as

$$S = \left[\begin{array}{cccc} x_1(1) & x_2(1) & \cdots & x_n(1) \\ x_1(1)p_1 & x_2(1)p_2 & \cdots & x_n(1)p_n \\ \vdots & \vdots & & \vdots \\ x_1(1)p_1^{m-(L+1)} & x_2(1)p_2^{m-(L+1)} & \cdots & x_n(1)p_n^{m-(L+1)} \\ \hline x_1(2) & x_2(2) & \cdots & x_n(2) \\ x_1(2)p_1 & x_2(2)p_2 & \cdots & x_n(2)p_n \\ \vdots & \vdots & & \vdots \\ x_1(2)p_1^{m-(L+1)} & x_2(2)p_2^{m-(L+1)} & \cdots & x_n(2)p_n^{m-(L+1)} \\ \hline & & \vdots & \\ \hline x_1(N) & x_2(N) & \cdots & x_n(N) \\ x_1(N)p_1 & x_2(N)p_2 & \cdots & x_n(N)p_n \\ \vdots & \vdots & & \vdots \\ x_1(N)p_1^{m-(L+1)} & x_2(N)p_2^{m-(L+1)} & \cdots & x_n(N)p_n^{m-(L+1)} \end{array}\right] G, \tag{31}$$

or

$$S = HG, \tag{32}$$

where G is given by

$$G = \left[\begin{array}{cccc} p_1 & p_1^2 & \cdots & p_1^L \\ p_2 & p_2^2 & \cdots & p_2^L \\ \vdots & \vdots & & \vdots \\ p_n & p_n^2 & \cdots & p_n^L \end{array}\right]. \tag{33}$$

Equation (30) thus becomes

$$HG\tilde{b} = -HGS^{+}\epsilon. \tag{34}$$

Now note from Equation (8) that the true and estimated Lth order characteristic polynomials are $B(z) = 1 + b_1 z^1 + b_2 z^2 + \cdots + b_L z^L$ and $\widehat{B}(z) = 1 + \hat{b}_1 z^1 + \hat{b}_2 z^2 + \cdots + \hat{b}_L z^L$, respectively. Note that $B(p_i) = 0$ and $\widehat{B}(\hat{p}_i) = 0$.

We can use a first order Taylor expansion to find an expression for the error in the estimated pole locations. We follow the technique in (Stoica *et al.*, 1989). For each $\hat{p}_i$ we obtain

$$0 \;=\; \widehat{B}(\hat{p}_i)$$

$$
\begin{aligned}
&= \widehat{B}(p_i) + \frac{\partial}{\partial z}\widehat{B}(z)|_{z=p_i}(\widehat{p}_i - p_i) + (\text{higher order terms}) \\
&= \widehat{B}(p_i) - B(p_i) + \frac{\partial}{\partial z}\widehat{B}(z)|_{z=p_i}(\widehat{p}_i - p_i) + (\text{higher order terms}) \\
&\approx 1 + \widehat{b}_1 p_i + \widehat{b}_2 p_i^2 + \cdots + \widehat{b}_L p_i^L - \left(1 + b_1 p_i + b_2 p_i^2 + \cdots + b_L p_i^L\right) \\
&\quad + \left(\widehat{b}_1 + 2\widehat{b}_2 p_i + \cdots + L\widehat{b}_L p_i^{L-1}\right)(\widehat{p}_i - p_i) \\
&\approx \begin{bmatrix} p_i & p_i^2 & \cdots & p_i^L \end{bmatrix} \begin{bmatrix} \widehat{b}_1 - b_1 \\ \widehat{b}_2 - b_2 \\ \vdots \\ \widehat{b}_L - b_L \end{bmatrix} + \begin{bmatrix} b_1 & b_2 & \cdots & b_L \end{bmatrix} \begin{bmatrix} 1 \\ 2p_i \\ \vdots \\ Lp_i^{L-1} \end{bmatrix} (\widehat{p}_i - p_i) \\
&= \begin{bmatrix} p_i & p_i^2 & \cdots & p_i^L \end{bmatrix} \left(\widehat{b} - b\right) + \eta_i (\widehat{p}_i - p_i). \qquad (35)
\end{aligned}
$$

Thus, to a first-order approximation

$$
(\widehat{p}_i - p_i) = -\frac{1}{\eta_i} \begin{bmatrix} p_i & p_i^2 & \cdots & p_i^L \end{bmatrix} \widetilde{b}, \qquad (36)
$$

where η_i is given by

$$
\eta_i = \begin{bmatrix} b_1 & b_2 & \cdots & b_L \end{bmatrix} \begin{bmatrix} 1 \\ 2p_i \\ \vdots \\ Lp_i^{L-1} \end{bmatrix}. \qquad (37)
$$

In matrix form we thus obtain

$$
\widetilde{P} \stackrel{\triangle}{=} \widehat{P} - P = -FG\widetilde{b}, \qquad (38)
$$

where

$$
\begin{aligned}
\widehat{P} &= \begin{bmatrix} \widehat{p}_1 & \widehat{p}_2 & \cdots & \widehat{p}_n \end{bmatrix}^T \\
P &= \begin{bmatrix} p_1 & p_2 & \cdots & p_n \end{bmatrix}^T \\
F &= \operatorname{diag}\left(\frac{1}{\eta_1}, \frac{1}{\eta_2}, \ldots, \frac{1}{\eta_n}\right). \qquad (39)
\end{aligned}
$$

Since the n poles are distinct, H is of full column rank. Hence, we can multiply Equa-

tion (34) by $(H^*H)^{-1}H^*$ to get

$$G\widetilde{b} = -GS^+\epsilon, \tag{40}$$

and by substituting Equation (40) into Equation (38) we obtain

$$\widetilde{P} = FGS^+\epsilon. \tag{41}$$

We now note that to a first order approximation, $\widetilde{P}$ is given by

$$\begin{aligned}
\widetilde{P} &= \widehat{\alpha} \odot e^{j\widehat{\omega}} - \alpha \odot e^{j\omega} \\
&= (\alpha + \widetilde{\alpha}) \odot e^{j(\omega+\widetilde{\omega})} - \alpha \odot e^{j\omega} \\
&= (\alpha + \widetilde{\alpha}) \odot e^{j\omega} \odot (1 + j\widetilde{\omega} + (\text{h.o.t.})) - \alpha \odot e^{j\omega} \\
&\approx (\alpha + \widetilde{\alpha}) \odot e^{j\omega} + \alpha \odot e^{j\omega} \odot j\widetilde{\omega} - \alpha \odot e^{j\omega} \\
&= T_p^{-1}\left(T_\alpha \widetilde{\alpha} + j\widetilde{\omega}\right),
\end{aligned} \tag{42}$$

where $\odot$ denotes the Hadamard product, and

$$\begin{aligned}
e^{j\omega} &= \begin{bmatrix} e^{j\omega_1} & e^{j\omega_2} & \cdots & e^{j\omega_n} \end{bmatrix}^T \\
\widetilde{\omega} &= \begin{bmatrix} \widetilde{\omega}_1 & \widetilde{\omega}_2 & \cdots & \widetilde{\omega}_n \end{bmatrix}^T \\
\widetilde{\alpha} &= \begin{bmatrix} \widetilde{\alpha}_1 & \widetilde{\alpha}_2 & \cdots & \widetilde{\alpha}_n \end{bmatrix}^T \\
T_p &= \operatorname{diag}\left(\frac{1}{p_1}, \frac{1}{p_2}, \ldots, \frac{1}{p_n}\right)
\end{aligned} \tag{43}$$

and where $\widetilde{\omega}_i = \widehat{\omega}_i - \omega_i$ and $\widetilde{\alpha}_i = \widehat{\alpha}_i - \alpha_i$. From Equations (41) and (42) we obtain:

$$\begin{aligned}
\widetilde{\omega} &= \operatorname{Im}\left\{T_p FGS^+\epsilon\right\} \\
\widetilde{\alpha} &= \operatorname{Re}\left\{T_\alpha^{-1} T_p FGS^+\epsilon\right\}.
\end{aligned} \tag{44}$$

Recall that the elements of W and w are uncorrelated, zero mean, complex white Gaussian random variables. Thus, ϵ is multivariate Gaussian with zero mean and covariance matrix

$$E\left[\epsilon\epsilon^*\right] = E\left[\left(\begin{bmatrix} w & : & W \end{bmatrix}\begin{bmatrix} 1 \\ b \end{bmatrix}\right)\left(\begin{bmatrix} w & : & W \end{bmatrix}\begin{bmatrix} 1 \\ b \end{bmatrix}\right)^*\right] = \sigma DD^*. \tag{45}$$

D is defined as a $(m-L)\,N \times mN$ block diagonal matrix given by

$$D = \text{diag}\,(B, B, \ldots, B)\,, \tag{46}$$

where B is given by

$$B = \begin{bmatrix} 1 & b_1 & b_2 & \cdots & b_L & 0 & 0 & \cdots & 0 \\ 0 & 1 & b_1 & \cdots & b_{L-1} & b_L & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 1 & b_1 & \cdots & b_{L-1} & b_L & 0 \\ 0 & \cdots & 0 & 0 & 1 & \cdots & b_{L-2} & b_{L-1} & b_L \end{bmatrix}_{(m-L)\times(m)} . \tag{47}$$

We also have

$$E\left[\epsilon\epsilon^T\right] = E\left[\left(\begin{bmatrix} w & : & W \end{bmatrix}\begin{bmatrix} 1 \\ b \end{bmatrix}\right)\left(\begin{bmatrix} w & : & W \end{bmatrix}\begin{bmatrix} 1 \\ b \end{bmatrix}\right)^T\right] = 0. \tag{48}$$

Using these results, along with the following relationships (proven in Stoica and Nehorai (1989))

$$\begin{aligned} \text{Re}\{u\}\text{Re}\{v^T\} &= \frac{1}{2}\left[\text{Re}\{uv^T\} + \text{Re}\{uv^*\}\right] \\ \text{Im}\{u\}\text{Im}\{v^T\} &= -\frac{1}{2}\left[\text{Re}\{uv^T\} - \text{Re}\{uv^*\}\right] \\ \text{Re}\{u\}\text{Im}\{v^T\} &= \frac{1}{2}\left[\text{Im}\{uv^T\} - \text{Im}\{uv^*\}\right], \end{aligned} \tag{49}$$

we obtain the following covariances for the pole parameters:

$$\begin{aligned} E\left[\widetilde{\omega}\widetilde{\omega}^T\right] &= \frac{\sigma}{2}\text{Re}\left\{T_p F G S^+ D D^* S^{*+} G^* F^* T_p^*\right\} \\ E\left[\widetilde{\omega}\widetilde{\alpha}^T\right] &= \frac{\sigma}{2}\text{Im}\left\{T_p F G S^+ D D^* S^{*+} G^* F^* T_p^* T_\alpha^{*-1}\right\} \\ E\left[\widetilde{\alpha}\widetilde{\alpha}^T\right] &= \frac{\sigma}{2}\text{Re}\left\{T_\alpha^{-1} T_p F G S^+ D D^* S^{*+} G^* F^* T_p^* T_\alpha^{*-1}\right\}. \end{aligned} \tag{50}$$

To obtain the covariances for the amplitude coefficient parameters we use Equation (14), which provides the amplitude coefficient estimates for each snapshot in terms of the estimated poles. We now note the following

$$\widetilde{X} \triangleq \widehat{X} - X \;=\; \widehat{A}^+ Y_a - A^+ S_a$$

$$
\begin{aligned}
&= \widehat{A}^{+}\left(S_a + N_a\right) - A^{+}S_a \\
&= \left(\widehat{A}^{+} - A^{+}\right) S_a + \widehat{A}^{+} N_a, \qquad (51)
\end{aligned}
$$

where

$$
X = \left[\begin{array}{cccc} x(1) & x(2) & \cdots & x(N) \end{array} \right], \qquad (52)
$$

and where S_a is the noise free version of Y_a and $N_a = \left[\begin{array}{cccc} e(1) & e(2) & \cdots & e(N) \end{array} \right]$ is the corresponding noise.

We apply the identity in Equation (25) to the first term of Equation (51). Since the n poles are distinct, A and $\widehat{A}$ are full rank. Also, since $m > n$ and $S_a \in \text{Range}(A)$, we have

$$
(\widehat{A}^{+} - A^{+})S_a = -\widehat{A}^{+}\widetilde{A}A^{+}S_a. \qquad (53)
$$

From Equations (51) and (53) we obtain the following first order approximation

$$
\begin{aligned}
\widetilde{X} &= -\widehat{A}^{+}\widetilde{A}A^{+}S_a + \widehat{A}^{+}N_a \\
&= -\widehat{A}^{+}\widetilde{A}X + \widehat{A}^{+}N_a \\
&\approx -A^{+}\widetilde{A}X + A^{+}N_a. \qquad (54)
\end{aligned}
$$

Note that $\widetilde{X}$ is a matrix in which each column is composed of the amplitude coefficient variations for each snapshot:

$$
\widetilde{X} = \left[\begin{array}{cccc} \widetilde{x}(1) & \widetilde{x}(2) & \cdots & \widetilde{x}(N) \end{array} \right]. \qquad (55)
$$

Following the same procedure as in Equation (42), a first order approximation of $\widetilde{x}(t)$ is given by

$$
\widetilde{x}(t) \approx T_x^{-1}(t)\left(T_\beta(t)\widetilde{\beta}(t) + j\widetilde{\gamma}(t)\right), \qquad (56)
$$

where

$$
\begin{aligned}
\widetilde{\gamma}(t) &= \left[\begin{array}{cccc} \widetilde{\gamma}_1(t) & \widetilde{\gamma}_2(t) & \cdots & \widetilde{\gamma}_n(t) \end{array} \right]^T \\
\widetilde{\beta}(t) &= \left[\begin{array}{cccc} \widetilde{\beta}_1(t) & \widetilde{\beta}_2(t) & \cdots & \widetilde{\beta}_n(t) \end{array} \right]^T \\
T_x(t) &= \text{diag}\left(\frac{1}{x_1(t)}, \frac{1}{x_2(t)}, \ldots, \frac{1}{x_n(t)}\right) \qquad (57)
\end{aligned}
$$

and where $\widetilde{\gamma}_i(t) = \widehat{\gamma}_i(t) - \gamma_i(t)$ and $\widetilde{\beta}_i(t) = \widehat{\beta}_i(t) - \beta_i(t)$. From Equations (54), (55), and (56) we obtain

$$\begin{aligned}
\widetilde{\gamma}(t) &= \mathrm{Im}\left\{T_x(t)A^+\left(-\widetilde{A}x(t) + e(t)\right)\right\} \\
\widetilde{\beta}(t) &= \mathrm{Re}\left\{T_\beta^{-1}(t)T_x(t)A^+\left(-\widetilde{A}x(t) + e(t)\right)\right\}. \qquad (58)
\end{aligned}$$

Before computing covariances for the parameters in Equation (58), we need to perform some manipulations to $\widetilde{A}x(t)$, since the random variable $\widetilde{A}$ does not appear at the rightmost position. We proceed with a first order approximation as follows

$$\begin{aligned}
\widetilde{A}x(t) &= \begin{bmatrix} 0 & 0 & \cdots & 0 \\ \widehat{p}_1 - p_1 & \widehat{p}_2 - p_2 & \cdots & \widehat{p}_n - p_n \\ \widehat{p}_1^2 - p_1^2 & \widehat{p}_2^2 - p_2^2 & \cdots & \widehat{p}_n^2 - p_n^2 \\ \vdots & \vdots & & \vdots \\ \widehat{p}_1^{m-1} - p_1^{m-1} & \widehat{p}_2^{m-1} - p_2^{m-1} & \cdots & \widehat{p}_n^{m-1} - p_n^{m-1} \end{bmatrix} x(t) \\
&\approx \begin{bmatrix} 0 & 0 & \cdots & 0 \\ \widetilde{p}_1 & \widetilde{p}_2 & \cdots & \widetilde{p}_n \\ 2p_1\widetilde{p}_1 & 2p_2\widetilde{p}_2 & \cdots & 2p_n\widetilde{p}_n \\ \vdots & \vdots & & \vdots \\ (m-1)\,p_1^{m-2}\widetilde{p}_1 & (m-1)\,p_2^{m-2}\widetilde{p}_2 & \cdots & (m-1)\,p_n^{m-2}\widetilde{p}_n \end{bmatrix} x(t) \\
&= CAT_p\mathrm{diag}\,(\widetilde{p}_1, \widetilde{p}_2, \ldots, \widetilde{p}_n)\,x(t) \\
&= CAT_x^{-1}(t)T_p\widetilde{P} \\
&= CAT_x^{-1}(t)T_pFGS^+\epsilon, \qquad (59)
\end{aligned}$$

where C is a diagonal matrix given by

$$C = \mathrm{diag}\,(0, 1, \ldots, m-1)\,. \qquad (60)$$

Equation (58) can now be approximated by

$$\begin{aligned}
\widetilde{\gamma}(t) &\approx \mathrm{Im}\left\{T_x(t)A^+\left(-CAT_x^{-1}(t)T_pFGS^+\epsilon + e(t)\right)\right\} \\
\widetilde{\beta}(t) &\approx \mathrm{Re}\left\{T_\beta^{-1}(t)T_x(t)A^+\left(-CAT_x^{-1}(t)T_pFGS^+\epsilon + e(t)\right)\right\}, \qquad (61)
\end{aligned}$$

Since

$$E\left[\epsilon e^*(t)\right] = E\left[\left(\begin{bmatrix} w & W \end{bmatrix}\begin{bmatrix}1\\ b\end{bmatrix}\right) e^*(t)\right] = \sigma D(t), \tag{62}$$

where $D(t)$ are each given by the tth column block of D (*cf.* see Equation (46)), we also have

$$\begin{aligned} E\left[\epsilon e^T(t)\right] &= E\left[\left(\begin{bmatrix} w & W \end{bmatrix}\begin{bmatrix}1\\ b\end{bmatrix}\right) e^T(t)\right] = 0 \\ E\left[e(t)e^T(t)\right] &= 0. \end{aligned} \tag{63}$$

Using these results and Equations (49) and (61) we obtain the following covariances for the amplitude coefficient parameters

$$\begin{aligned} E\left[\widetilde{\gamma}(t)\widetilde{\gamma}^T(r)\right] &= \frac{\sigma}{2}\mathrm{Re}\Big\{T_x(t)A^+\Big(CAT_x^{-1}(t)T_pFGS^+DD^*S^{+*}G^*F^*T_p^*T_x^{*-1}(r)A^*C^* \\ &\quad -CAT_pT_x^{-1}(t)FGS^+D(r) - D^*(t)S^{+*}G^*F^*T_p^*T_x^{*-1}(r)A^*C^* \\ &\quad +I_m\delta_{t,r}\Big)A^{+*}T_x^*(r)\Big\} \\ E\left[\widetilde{\gamma}(t)\widetilde{\beta}^T(r)\right] &= \frac{\sigma}{2}\mathrm{Im}\Big\{T_x(t)A^+\Big(CAT_x^{-1}(t)T_pFGS^+DD^*S^{+*}G^*F^*T_p^*T_x^{*-1}(r)A^*C^* \\ &\quad -CAT_x^{-1}(t)T_pFGS^+D(r) - D^*(t)S^{+*}G^*F^*T_p^*T_x^{*-1}(r)A^*C^* \\ &\quad +I_m\delta_{t,r}\Big)A^{+*}T_x^*(r)T_\beta^{-1*}(r)\Big\} \\ E\left[\widetilde{\beta}(t)\widetilde{\beta}^T(r)\right] &= \frac{\sigma}{2}\mathrm{Re}\Big\{T_\beta^{-1}(t)T_x(t)A^+\Big(CAT_x^{-1}(t)T_pFGS^+DD^*S^{+*}G^*F^*T_p^*T_x^{*-1}(r)A^*C^* \\ &\quad -CAT_x^{-1}(t)T_pFGS^+D(r) - D^*(t)S^{+*}G^*F^*T_p^*T_x^{*-1}(r)A^*C^* \\ &\quad +I_m\delta_{t,r}\Big)A^{+*}T_x^*(r)T_\beta^{-1*}(r)\Big\}. \end{aligned} \tag{64}$$

Using Equations (44), (49), and (61) we can also compute the cross-covariances between the poles and the amplitude coefficients as follows

$$\begin{aligned} E\left[\widetilde{\gamma}(t)\widetilde{\omega}^T\right] &= \frac{\sigma}{2}\mathrm{Re}\Big\{T_x(t)A^+\Big(-CAT_x^{-1}(t)T_pFGS^+DD^* + D^*(t)\Big)S^{+*}G^*F^*T_p^*\Big\} \\ E\left[\widetilde{\gamma}(t)\widetilde{\alpha}^T\right] &= \frac{\sigma}{2}\mathrm{Im}\Big\{T_x(t)A^+\Big(-CAT_x^{-1}(t)T_pFGS^+DD^* + D^*(t)\Big)S^{+*}G^*F^*T_p^*T_\alpha^{-1*}\Big\} \\ E\left[\widetilde{\beta}(t)\widetilde{\omega}^T\right] &= -\frac{\sigma}{2}\mathrm{Im}\Big\{T_\beta^{-1}(t)T_x(t)A^+\Big(-CAT_x^{-1}(t)T_pFGS^+DD^* + D^*(t)\Big)S^{+*}G^*F^*T_p^*\Big\} \\ E\left[\widetilde{\beta}(t)\widetilde{\alpha}^T\right] &= \frac{\sigma}{2}\mathrm{Re}\Big\{T_\beta^{-1}(t)T_x(t)A^+\Big(-CAT_x^{-1}(t)T_pFGS^+DD^* + D^*(t)\Big)S^{+*}G^*F^*T_p^*T_\alpha^{-1*}\Big\}. \end{aligned} \tag{65}$$

Equations (50), (64), and (65) completely specify Σ_θ as given in Equation (20) using the following substitutions

$$\begin{aligned}
U(t,r) &= R(t)ZR^*(r) - R(t)Q(r) - Q^*(t)R^*(r) + \frac{\sigma}{2}T_x(t)A^+A^{+*}T_x^*(r)\delta_{t,r} \\
V(t) &= -R(t)Z + Q^*(t) \\
Z &= \frac{\sigma}{2}T_pFGS^+DD^*S^{*+}G^*F^*T_p^* \\
Q(t) &= \frac{\sigma}{2}T_pFGS^+D(t)A^{+*}T_x^*(t) \\
R(t) &= T_x(t)A^+CAT_x^{-1}(t).
\end{aligned} \tag{66}$$

Furthermore, since ϵ and $e(t)$ are zero mean, Equations (44) and (61) imply that the parameter estimates are unbiased, *i.e.* to a first order approximation $E\left[\widehat{\theta}\right] = \theta$. This completes the proof. □

Table 1: Theoretical and simulation variances for the poles of a general ten mode case.

Pole number	Theoretical variance (dB)	Simulation variance (dB)	Pole radius
1	-24.2	-22.4	1.115
2	-42.7	-43.4	1.05
3	-35.7	-35.2	0.8
4	-25.8	-22.4	1.12
5	-57.1	-56.9	1.0
6	-57.0	-56.1	1.0
7	-25.8	-21.5	0.6
8	-29.4	-29.1	1.115
9	-39.1	-32.2	0.9
10	-30.1	-26.7	0.7

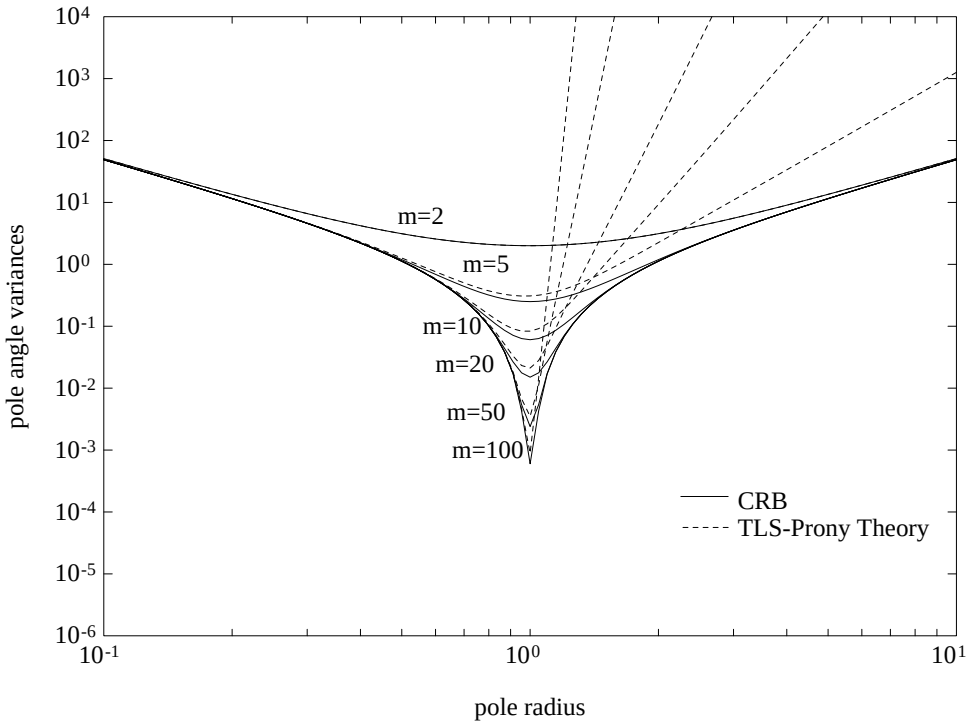


Figure 1: Pole angle variances for single pole data ($n = 1$).

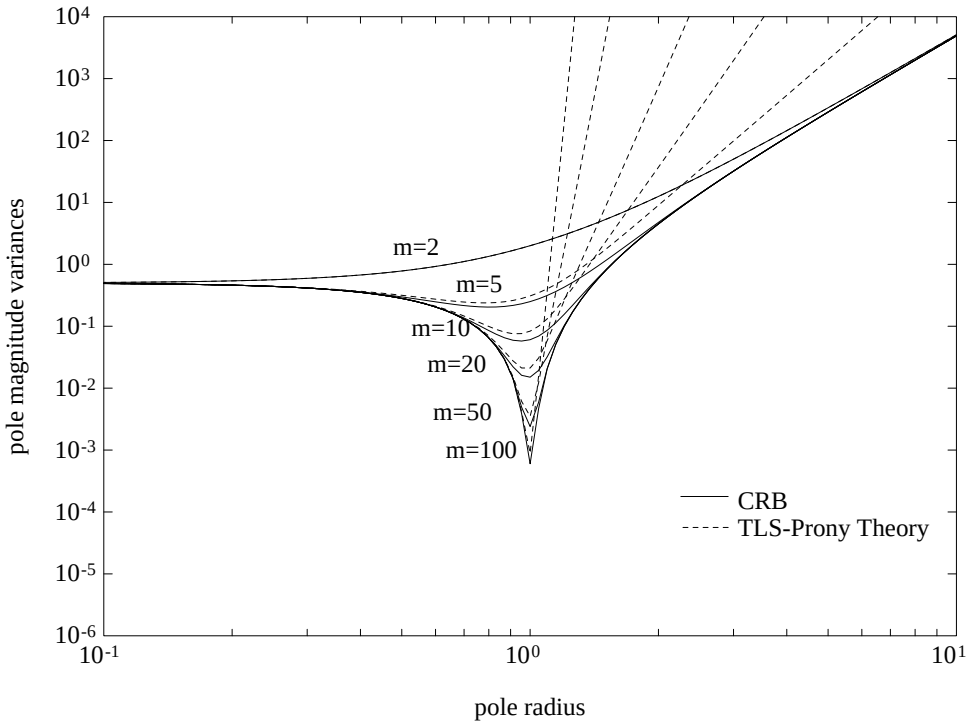


Figure 2: Pole magnitude variances for single pole data ($n = 1$).

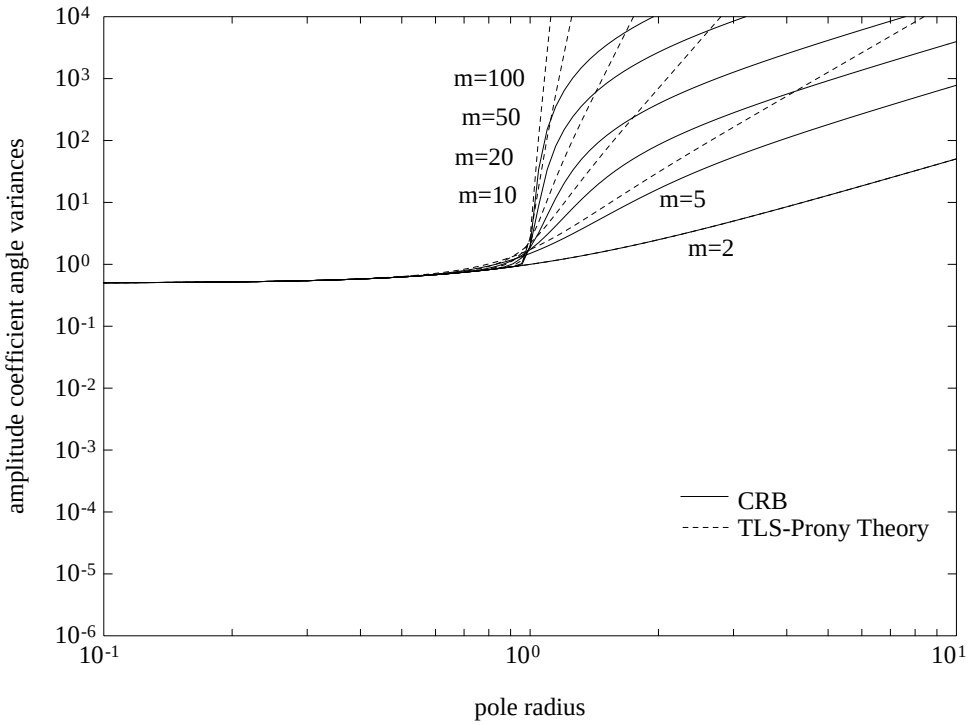


Figure 3: Amplitude coefficient angle variances for single pole data (n=1).

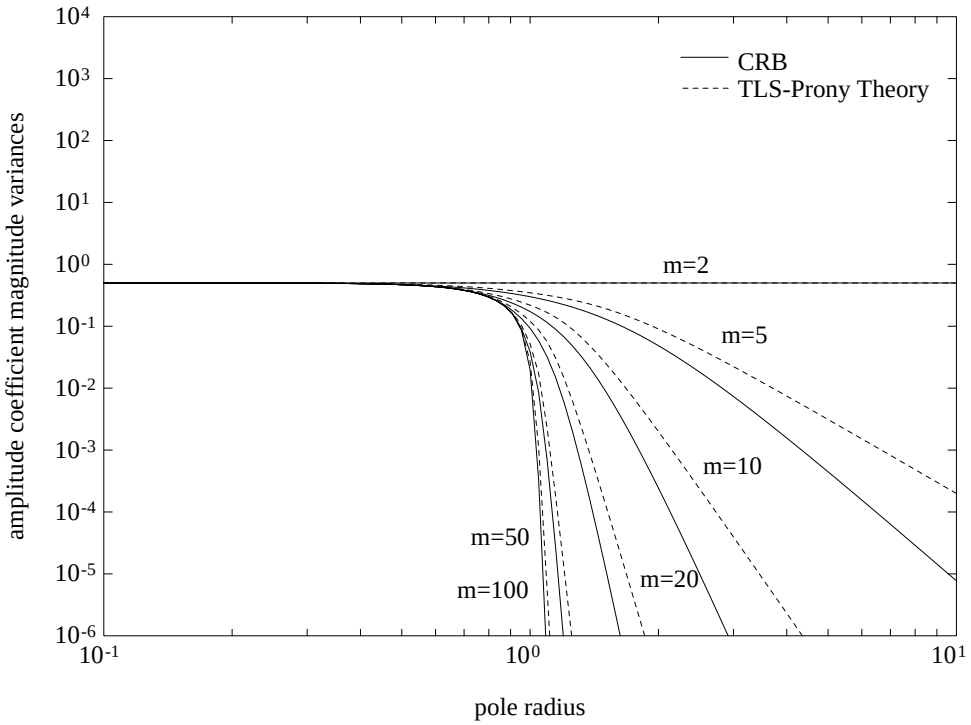


Figure 4: Amplitude coefficient magnitude variances for single pole data (n=1).

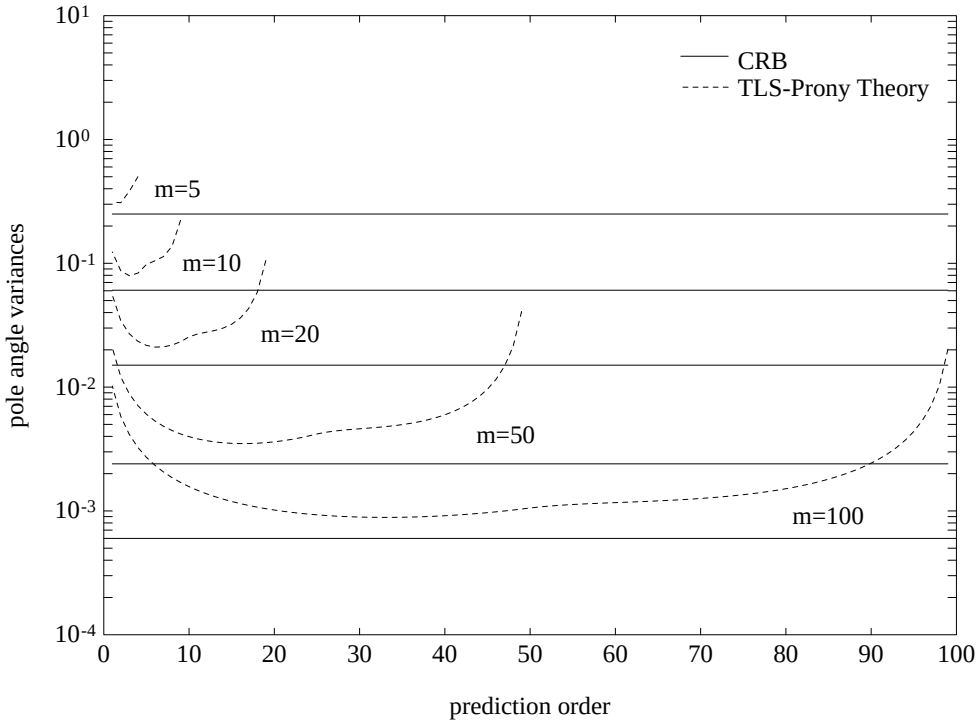


Figure 5: Pole angle variances for various prediction orders.

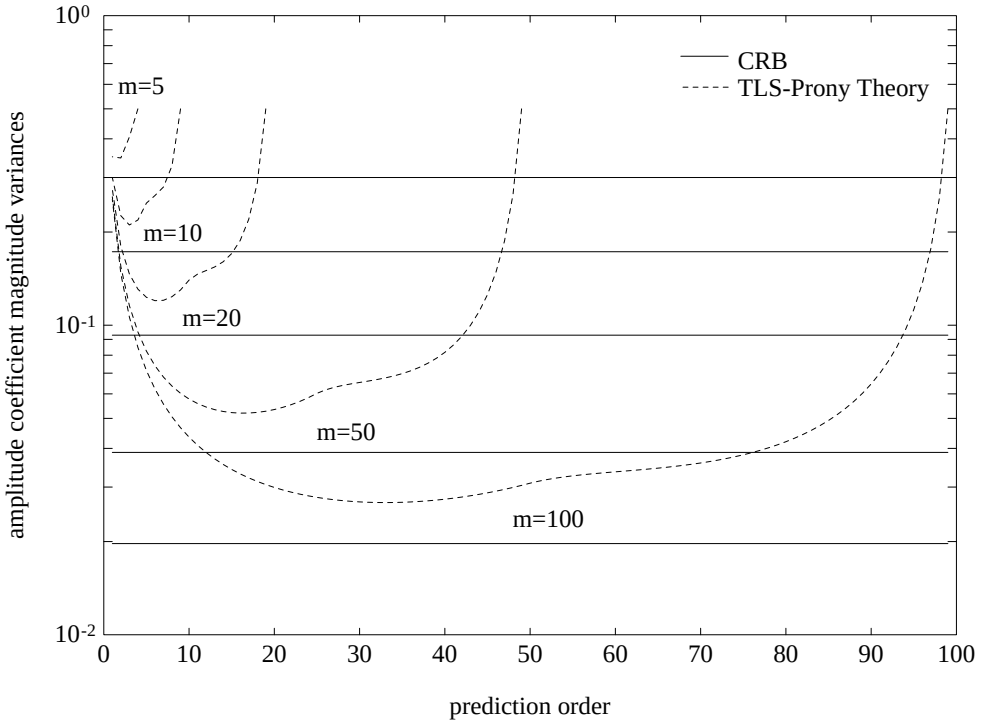


Figure 6: Amplitude coefficient magnitude variances for various prediction orders.

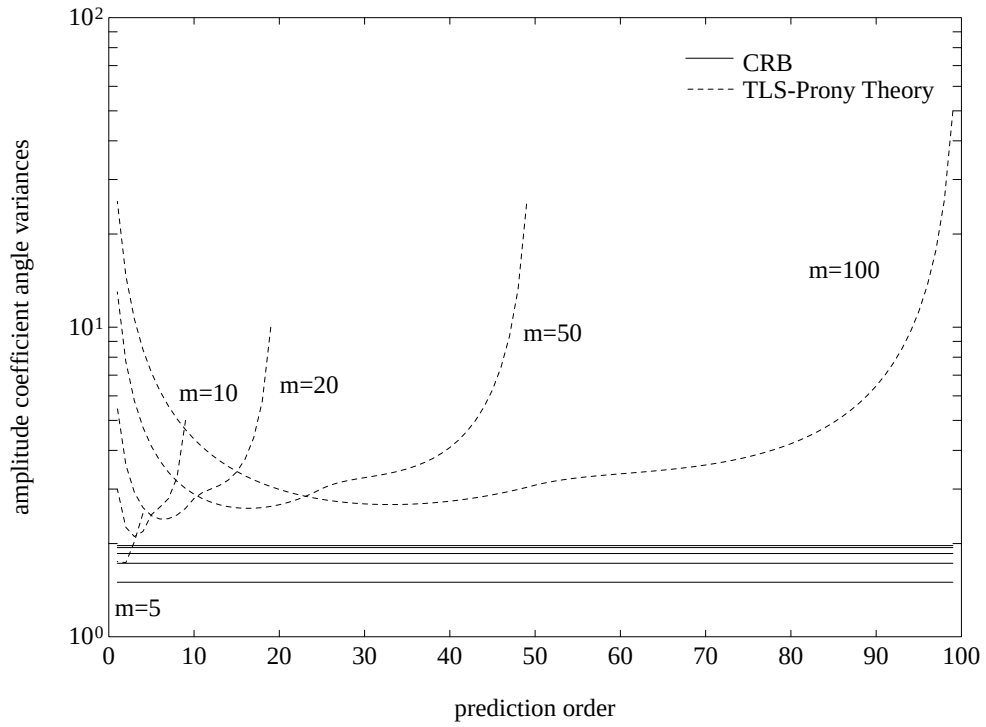


Figure 7: Amplitude coefficient angle variances for various prediction orders.

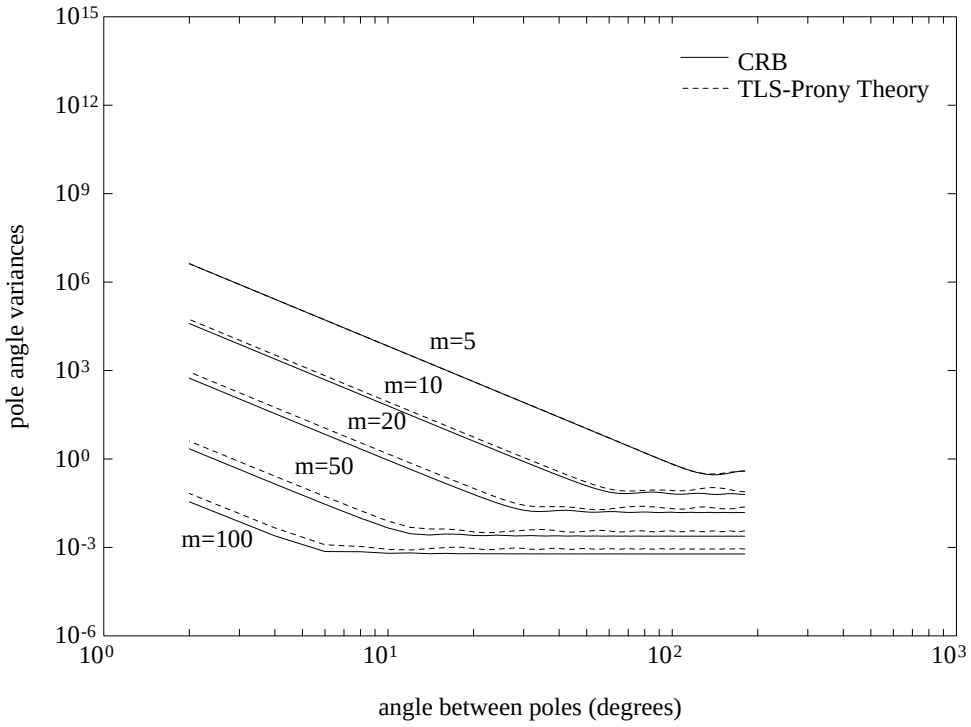
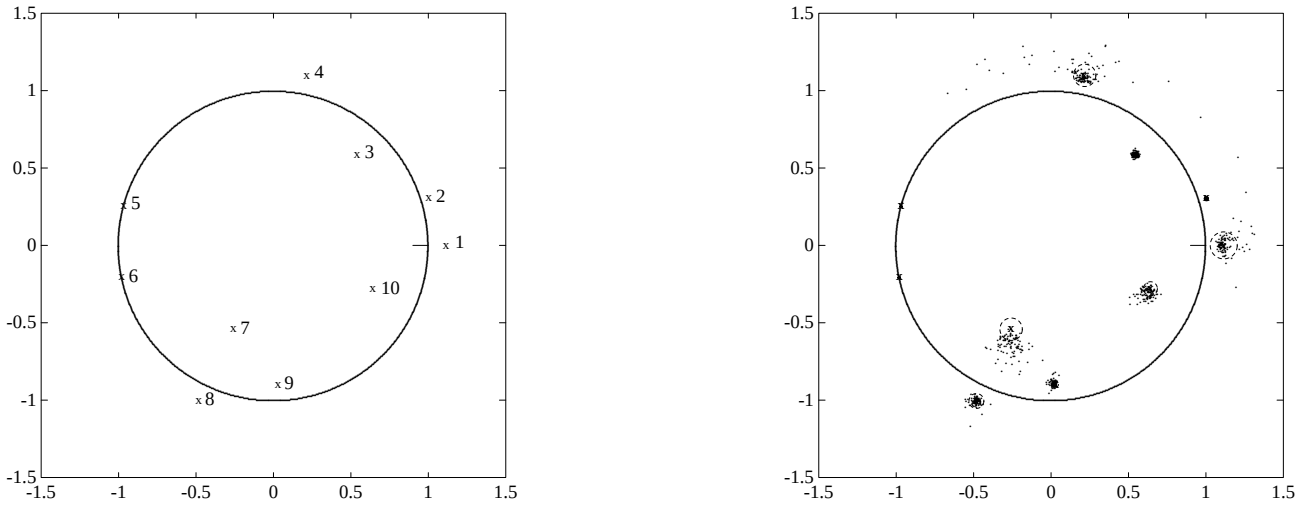


Figure 8: Angle variances as a function of pole separation for two poles on the unit circle.



(a) (b)

Figure 9: A general ten mode case: (a) true pole locations and (b) two standard deviation theoretical bounding circles and Monte-Carlo estimates.