STATISTICAL ANALYSIS OF SVD-BASED PRONY TECHNIQUES *

William M. Steedly **Ching-Hui J. Ying**
Randolph L. Moses
Department of Electrical Engineering
The Ohio State University
Columbus, Ohio 43210

Abstract

In this paper we present an analysis of parameter variance statistics for the SVD-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.

1 Introduction

The problem of estimating model parameters of noisy exponential signals is an active area of research. These models have a single set of complex poles and one or more set of amplitude coefficients (snapshots). The performance of these parameter estimation methods is often measured by the accuracy of the estimated poles, since these pole locations contain such information as formant frequencies or directions of arrival of signal components.

One of the popular estimation schemes is the SVD-Prony estimation [1]. The multi-snapshot generalization is considered in this paper. This paper provides a complete statistical derivation for the multi-snapshot case where signals consist of arbitrary exponential terms in noise. We have derived its asymptotic statistics, including the covariance matrix of the estimated parameters. The parameters include the magnitudes and angles of the poles, and the magnitudes and angles of the amplitude coefficients.

Using these expressions, it is shown that the angle and magnitude parameters are uncorrelated. It is also shown that if 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. In addition, we show that the transition from the $1/m^3$ (where m is the number of data points) variance bounds decrease for poles on the unit circle to the variance decrease as poles move off the unit circle is also detailed.

The effects on poles inside and outside the unit circle using backward or forward linear prediction in the SVD-Prony estimation scheme is also detailed.

2 Estimation Procedure Review

2.1 Data Model

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

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

Each data vector is modeled as a noisy exponential sequence

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

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.

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

Equation 2 may be compactly written as

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

where $e(t) = \left[\begin{array}{cccc} e_0(t) & e_1(t) & \cdots & e_{m-1}(t) \end{array}\right]^T$, $x(t) = \left[\begin{array}{cccc} x_0(t) & x_1(t) & \cdots & x_{n-1}(t) \end{array}\right]^T$, and A is the $m \times n$ Vandermonde matrix derived from n signal poles

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

2.2 Parameter Estimation

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

$$Yb \approx -y, \tag{5}$$

where

$$b = \left[\begin{array}{cccc} b_1 & b_2 & \cdots & b_L \end{array}\right]^T \tag{6}$$

and where

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

Here L is the order of prediction; 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; in practice, choosing $L > n$ results in more accurate parameter estimates. 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 solution of Equation 5 involves obtaining a singular value decomposition of the matrix $\left[\begin{array}{cc} y & Y \end{array}\right]$ and truncating all but the first n singular values to arrive at a noise cleaned estimate $\left[\begin{array}{cc} \hat{y} & \hat{Y} \end{array}\right]$ [1]. This leads to the modified linear prediction equation

$$\hat{Y}\hat{b} = \hat{y} \tag{9}$$

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

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

where $^+$ denotes the Moore-Penrose pseudoinverse. Finally, the estimates for the poles are found by

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

Once the poles have been determined, the equation for the amplitude coefficients can be formed. Equation 3 leads to the following equation for the amplitude coefficients,

$$\begin{bmatrix} 1 & \cdots & 1 \\ \hat{p}_1 & \cdots & \hat{p}_L \\ \vdots & & \vdots \\ \hat{p}_1^{m-1} & \cdots & \hat{p}_L^{m-1} \end{bmatrix} \begin{bmatrix} \hat{x}_1(1) & \cdots & \hat{x}_1(N) \\ \vdots & & \vdots \\ \hat{x}_L(1) & \cdots & \hat{x}_L(N) \end{bmatrix} = \left[\begin{array}{ccc} y(1) & \cdots & y(N) \end{array}\right] \tag{12}$$

or

$$\hat{A}\hat{X} = Y_a. \tag{13}$$

The amplitude coefficients can be found from a least squares solution to Equation 13,

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

where * denotes complex conjugate transpose. Because only n singular values of $\hat{Y}$ are nonzero, there are at most n pole estimates which can correspond to true data modes. Therefore, only the n poles which have the largest energy are retained. Thus estimates $\hat{A}$ and $\hat{X}$ become matrices of sizes $m \times n$ and $n \times N$, where the $L - n$ columns and rows, respectively, corresponding to the lowest energy modes are deleted.

3 Statistical Analysis

To analyze parameter statistics we now derive their covariance matrix. This is given in the following theorem.

Theorem: Assume data is given by Equation 2. We define ω_i and α_i, to be the angle and magnitude, respectively, of each pole p_i, thus $p_i = \alpha_i e^{j\omega_i}$. Similarly we define $\gamma(t)$ and $\beta(t)$ to be the angle and magnitude vectors, respectively, of each vector of amplitude coefficients $x(t)$. We now define the following parameter vectors:

$$\begin{aligned} \theta_x &= \left[\begin{array}{cccc} \gamma^T(1) & \beta^T(1) & \cdots & \gamma^T(N) \quad \beta^T(N) \end{array}\right]^T \\ \theta_p &= \left[\begin{array}{cccc} \omega_1 & \cdots & \omega_n \quad \alpha_1 & \cdots \quad \alpha_n \end{array}\right]^T \\ \theta &= \left[\begin{array}{cc} \theta_x^T & \theta_p^T \end{array}\right]^T, \end{aligned} \tag{15}$$

and let $\widehat{\theta}$ denote the SVD-Prony estimate of θ, which are given by the n highest energy mode estimates found in Equations 11 and 14.

Then the asymptotic (high SNR, where SNR is defined as total signal power divided by total noise power) pdf of $\widehat{\theta}$ is given by

$$\widehat{\theta} \sim N(\theta, \Sigma_\theta), \tag{16}$$

where

$$\Sigma_\theta = \begin{bmatrix} \overline{U}(1,1) & \widetilde{U}(1,1)T_\beta^{*-1}(1) & \cdots & \overline{U}(1,N) \\ -T_\beta^{-1}(1)\widetilde{U}(1,1) & T_\beta^{-1}(1)\overline{U}(1,1)T_\beta^{*-1}(1) & \cdots & -T_\beta^{-1}(1)\widetilde{U}(1,N) \\ \vdots & \vdots & \ddots & \vdots \\ \overline{U}(N,1) & \widetilde{U}(N,1)T_\beta^{*-1}(1) & \cdots & \overline{U}(N,N) \\ -T_\beta^{-1}(N)\widetilde{U}(N,1) & T_\beta^{-1}(N)\overline{U}(N,1)T_\beta^{*-1}(1) & \cdots & -T_\beta^{-1}(N)\widetilde{U}(N,N) \\ \overline{V^*}(1) & \widetilde{V^*}(1)T_\beta^{*-1}(1) & \cdots & \overline{V^*}(N) \\ -T_\alpha^{-1}\widetilde{V^*}(1) & T_\alpha^{-1}\overline{V^*}(1)T_\beta^{*-1}(1) & \cdots & -T_\alpha^{-1}\widetilde{V^*}(N) \end{bmatrix}$$

$$\begin{bmatrix} \widetilde{U}(1,N)T_\beta^{*-1}(N) & \overline{V}(1) & \widetilde{V}(1)T_\alpha^{*-1} \\ T_\beta^{-1}(1)\overline{U}(1,N)T_\beta^{*-1}(N) & -T_\beta^{-1}(1)\widetilde{V}(1) & T_\beta^{-1}(1)\overline{V}(1)T_\alpha^{*-1} \\ \vdots & \vdots & \vdots \\ \widetilde{U}(N,N)T_\beta^{*-1}(N) & \overline{V}(N) & \widetilde{V}(N)T_\alpha^{*-1} \\ T_\beta^{-1}(N)\overline{U}(N,N)T_\beta^{*-1}(N) & -T_\beta^{-1}(N)\widetilde{V}(N) & T_\beta^{-1}(N)\overline{V}(N)T_\alpha^{*-1} \\ \widetilde{V^*}(N)T_\beta^{*-1}(N) & \overline{Z} & \widetilde{Z}T_\alpha^{*-1} \\ T_\alpha^{-1}\overline{V^*}(N)T_\beta^{*-1}(N) & -T_\alpha^{-1}\widetilde{Z} & T_\alpha^{-1}\overline{Z}T_\alpha^{*-1} \end{bmatrix} \tag{17}$$

where $\bar{\cdot}$ and $\tilde{\cdot}$ in Equation 17 are real and imaginary part operators, respectively, and where

$$\begin{aligned} U(t,r) &= R(t)ZR^*(r) - R(t)Q(r) - Q^*(t)R^*(r) \\ &\quad + \frac{\sigma}{2}T_x(t)A^+A^{+*}T_x^*(r)\delta_{t,r} \\ V(t) &= -R(t)Z + Q^*(t) \\ Z &= \frac{\sigma}{2}T_pFGS^+DD^*S^{*+}G^*F^*T_p^* \\ Q(t) &= \frac{\sigma}{2}T_pFGS^+D(t)A^{+*}T_x^*(t) \\ R(t) &= T_x(t)A^+CAT_x^{-1}(t) \end{aligned} \tag{18}$$

Here, $T_x(t)$, $T_\beta(t)$, T_p, T_α, C, and F are diagonal matrices given by

$$\begin{aligned} T_x(t) &= \operatorname{diag}\left(\frac{1}{x_1(t)}, \frac{1}{x_2(t)}, \ldots, \frac{1}{x_n(t)}\right) \\ T_\beta(t) &= \operatorname{diag}\left(\frac{1}{\beta_1(t)}, \frac{1}{\beta_2(t)}, \ldots, \frac{1}{\beta_n(t)}\right) \\ T_p &= \operatorname{diag}\left(\frac{1}{p_1}, \frac{1}{p_2}, \ldots, \frac{1}{p_n}\right) \\ T_\alpha &= \operatorname{diag}\left(\frac{1}{\alpha_1}, \frac{1}{\alpha_2}, \ldots, \frac{1}{\alpha_n}\right) \\ C &= \operatorname{diag}(0, 1, \ldots, m-1) \\ F &= \operatorname{diag}\left(\frac{1}{\eta_1}, \frac{1}{\eta_2}, \ldots, \frac{1}{\eta_n}\right) \end{aligned} \tag{19}$$

and η_i and G are given by

$$\eta_i = \begin{bmatrix} b_1 & b_2 & \cdots & b_L \end{bmatrix} \begin{bmatrix} 1 \\ 2p_i \\ \vdots \\ Lp_i^{L-1} \end{bmatrix} \tag{20}$$

$$G = \begin{bmatrix} p_1 & p_1^2 & \cdots & p_1^L \\ p_2 & p_2^2 & \cdots & p_2^L \\ \vdots & \vdots & & \vdots \\ p_n & p_n^2 & \cdots & p_n^L \end{bmatrix}. \tag{21}$$

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

$$D = \operatorname{diag}(B, B, \ldots, B), \tag{22}$$

where B is $(m-L) \times m$ and given by

$$B = \begin{bmatrix} 1 & b_1 & b_2 & \cdots & b_L & 0 & 0 & \cdots & 0 \\ 0 & 1 & b_1 & \cdots & b_{L-1} & b_L & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 1 & b_1 & \cdots & b_{L-1} & b_L & 0 \\ 0 & \cdots & 0 & 0 & 1 & \cdots & b_{L-2} & b_{L-1} & b_L \end{bmatrix}. \tag{23}$$

The matrices $D(t)$ are each given by the tth column block of D and S is the noise free version of Y. □

Proof: See [2].

Corollary: The elements of $\gamma(t)$ are uncorrelated with the corresponding elements of $\beta(t)$ and the elements ω are uncorrelated with corresponding elements of α. □

Proof: Note that the corresponding covariances in Σ_θ are the diagonal elements of the product of the imaginary part of a Hermitian matrix and a diagonal matrix and are thus identically zero. □

We can also note that the angle variances are equal to the magnitude variances except for the transformation matrices $T_x(t)$ and T_p. Without these transformation matrices, Σ_θ would correspond to a relative magnitude parameterization, where magnitude relative to the true magnitude is estimated, rather than absolute magnitude. When real and imaginary parts of the amplitude coefficients and poles are considered as parameters, the resulting variances can be derived using a simple Jacobian coordinate transformation and are equal.

Another property which can be pointed out is absolute phase invariance. We can see that the terms in

Equation 18 remain unchanged as $T_x(t) \rightarrow e^{j\phi(t)}T_x(t)$ and $p_i \rightarrow e^{j\phi}p_i$ due to cancellations. Thus the covariance matrix depends only on relative phase between the poles and between each set of amplitude coefficients for each snapshot.

4 Simulation Studies

Simulations have been performed using the statistical analyses above. Recently a CRB formulation has been developed for multi-snapshot damped exponentials in noise [3]. These CRB results are compared with the variances of the estimated poles using the SVD-Prony method to examine its performance.

4.1 Variance Results for Single Exponential Case

In this simulation, a model with a single pole and with one snapshot of data was chosen. The experiment entailed moving the pole along the positive real axis (the results are independent of the pole angle, so an angle of zero was chosen) from 0.1 to 10 and calculating the variances using Equation 17 for data sets of lengths 2, 5, 10, 20, 50, and 100. For comparative purposes, the amplitude coefficient associated with the pole was chosen to be a positive real number such that the mode energy ($x^2 \sum_{l=0}^{m-1} p^{2l}$) was unity for each pole location and data length. The noise power was also kept constant at $\sigma = 1$. The model order was assumed to be one third of the data length, which has been shown to be optimal [4, 5].

The variances for the pole angle and 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 just above the CRBs inside and near the unit circle. Outside the unit circle the variances become much higher (except for the $m = 2$ case) than the CRBs. This is due to the fact that backward linear prediction is used in this SVD-Prony method; note that with backward linear prediction extraneous poles are placed outside the unit circle, thus making estimation of poles outside the unit circle more difficult. Using forward linear prediction will give the opposite results for poles inside and outside the unit circle. Similar observations apply to the pole magnitude variances (*cf.* Figure 2)

From these two figures we 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 due to the fact that angular

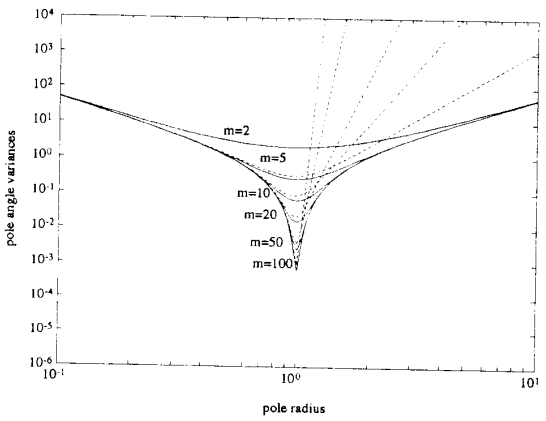

Figure 1: Pole angle variances for single pole data ($n = 1$).

uncertainty becomes greater as a pole moves closer to the origin. Note that the pole angle variance approaches infinity as the pole approaches the origin, which is what one would expect since its angle is undefined at the origin.

We see that the variances for both pole angle and magnitude are asymptotically (as $m \rightarrow \infty$) 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 was adjusted in this experiment to keep the mode energy constant (if the amplitude coefficient is left unchanged, the variance decrease is $1/m^3$). Note, however, that as the number of data points is reduced, the pole magnitude which gives the minimum variance is less than unity.

When the pole is not on the unit circle, the variances do not decrease to zero as $m \rightarrow \infty$. Because of the decay or growth of the exponential mode, and because the mode energy is kept constant, increasing m results in adding data points with lower and lower amplitude. Thus, the variances do not continue to decrease.

The variances for the amplitude coefficient angle and magnitude are also available and appear in [2].

4.2 Some General Observations

Using the covariance matrix, one can determine a confidence region corresponding to an estimate of each pole. This region can be plotted as, for example, a two standard deviation confidence "ellipse" about

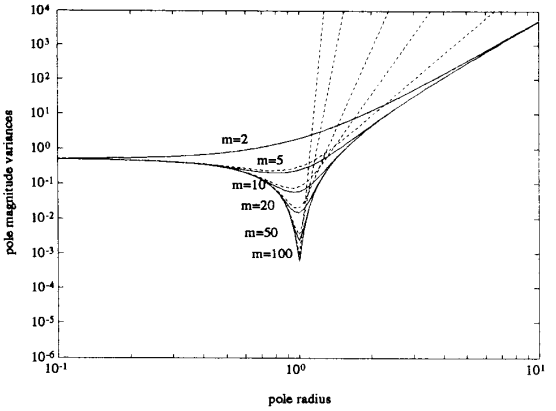

Figure 2: Pole magnitude variances for single pole data ($n = 1$).

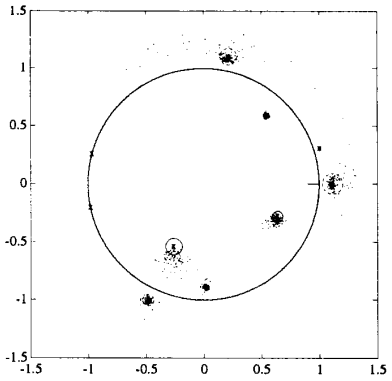

Figure 3: Two standard deviation bounding circles for each pole, using a tenth order model, with $m = 100$ data points, $\sigma = 0.001$, and each mode energy set to unity.

each pole. Figure 3 is an example of such a plot, where the pole locations are indicated by "x" symbols. In this experiment there are ten poles, each with an energy of one. Here, $m = 100$ data points and the noise power is $\sigma = 0.001$. This gives an SNR of 10dB per mode. The dashed ellipses (they are actually circles due to the corollary and the equivalence of the real and imaginary part variances as noted in Section 3) shown in Figure 3 represent the two standard deviation boundaries around the ten pole locations. 87% of poles found in a Monte-Carlo simulation would be expected to fall within these circles. Note that the boundary circles are significantly smaller for the poles which are located close to the unit circle. The solid circles represent the corresponding CRB bounds. One hundred Monte-Carlo simulations were performed and the estimated poles appear as dots in Figure 3. From these estimates, we can see that the statistical analysis is in general confirmed; the exception being some bias in the estimates of the pole closest to the origin.

5 Conclusions

We presented a statistical analysis for estimated poles of the SVD-Prony algorithm. We have provided complete expressions for the covariance matrix of the parameters of an exponential model with 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.

References

[1] 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*, vol. ASSP-35, pp. 1440–1454, Oct. 1987.

[2] W. M. Steedly, C. J. Ying, and R. L. Moses, "Statistical analysis of SVD-based Prony techniques," Submitted for publication.

[3] 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*, (Toronto, Ontario), pp. 3569–3572, May 14–17, 1991.

[4] 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*, vol. ASSP-38, pp. 814–824, May 1990.

[5] W. M. Steedly, C. J. Ying, and R. L. Moses, "A modified SVD-Prony method using data decimation for computationally efficient estimation of damped exponentials in noise," *IEEE Transactions on Signal Processing*, Submitted for publication.