A modified TLS-Prony method using data decimation

William M. Steedly, Chinghui J. Ying, and Randolph L. Moses

Department of Electrical Engineering
The Ohio State University
Columbus, Ohio 43210

ABSTRACT

This paper introduces a modified TLS-Prony method which uses data decimation. The use of data decimation results in the reduction in the computational complexity in a variety of applications by allowing several low order estimations to be performed rather than one high order estimation. We also present an analysis of pole variance statistics for the modified TLS-Prony method which is used to explain and quantify the characteristics of decimation. We show that using decimation we can obtain comparable performance results at a fraction of the computational cost.

1. INTRODUCTION

A popular high resolution estimation technique is the use of backward linear prediction coupled with singular value decomposition (SVD) and total least squares,[1] here called TLS-Prony. This technique has been shown to provide good parameter estimates of damped exponential signals in noise for various types of data.[1,2] This paper incorporates decimation into the TLS-Prony estimation scheme.

Decimation has been considered before in the context of spectral estimation.[3,4] This technique entails using only part of the measured data. Decimation of correlation sequences was also considered in [5]; this technique effectively uses all the measured data, but is somewhat restrictive in that it applies only to correlation-based parameter estimation techniques. Moreover, none of these works analyze the statistical properties of the resulting parameter estimates.

In this paper we present a data decimation technique which is based on the TLS-Prony algorithm.[1] We also develop a theoretical statistical analysis of the accuracy of the TLS-Prony parameter estimates when decimation is used. This analysis permits a quantitative comparison of estimation accuracy for various types of data decimation schemes. It also facilitates a comparison of the decimated and undecimated procedures in terms of estimation accuracy.

A decimation based estimator is motivated by the need for reduction in the computational complexity of estimators in a variety of applications. There are two cases of interest. First, the signals of interest may be bandlimited and occupy a relatively small region of the unambiguous frequency range $f \in \left[-\frac{1}{2}, \frac{1}{2}\right]$. In this case one is interested in analyzing a subset of the whole frequency range. For example, this technique was used to investigate radar signatures of aircraft.[2,6] One question is whether decimation can improve the accuracy of the pole estimates in this case.

A second case of interest is when the signal occupies most or all of the unambiguous frequency range. In this case we filter the data to isolate a number of subbands, then use a decimation version of TLS-Prony to estimate the poles in each of the subbands. This idea is similar in principle to beamspace prefiltering in array processing.[7,8,9] By focusing on particular bands one at a time, estimation techniques can be used with lower model orders since there are typically fewer modes within each of the bands. Thus, a single wideband estimation procedure is replaced by several lower order estimations. This has the advantages of being much less numerically intensive and of being amenable to parallel implementation.[9]

In this paper an TLS-Prony decimation estimation procedure is presented. Statistical analysis results for this procedure are derived and compared to the Cramér-Rao bound (CRB). The statistical analysis demonstrates that the

performance of the decimated TLS-Prony procedure is comparable to the performance of the undecimated TLS-Prony procedure for undamped exponential modes. A complexity analysis is also included and shows that the decimated algorithm is computationally more efficient than the undecimated algorithm. Both the statistical performance and the decrease in computational complexity are verified by Monte-Carlo simulations.

2. DECIMATION ESTIMATION PROCEDURE

2.1 Data Model

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

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

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 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

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

2.2 Parameter Estimation

In data decimation, the data is filtered to a band no larger than $\frac{1}{d}$ so that aliasing does not occur (see discussion in Section 4). We will denote this filtered data as $y_q'(t)$. This filtered data contains only $n' \leq n$ modes. After filtering this data can be decimated by a factor of d.

The backward linear prediction equations use the decimated data. Thus the decimated 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 $b = \begin{bmatrix} b_1 & b_2 & \cdots & b_L \end{bmatrix}^T$ and

$$\begin{bmatrix} y' & Y' \end{bmatrix} = \left[\begin{array}{cccc} y'(1) & y'_d(1) & y'_{2d}(1) & \cdots & y'_{Ld}(1) \\ y'_d(1) & y'_{2d}(1) & y'_{3d}(1) & \cdots & y'_{(L+1)d}(1) \\ \vdots & \vdots & \vdots & & \vdots \\ y'_{m-(L+1)d}(1) & y'_{m-Ld}(1) & y'_{m-(L-1)d}(1) & \cdots & y'_{m-d}(1) \\ \hline y'(2) & y'_d(2) & y'_{2d}(2) & \cdots & y'_{Ld}(2) \\ y'_d(2) & y'_{2d}(2) & y'_{3d}(2) & \cdots & y'_{(L+1)d}(2) \\ \vdots & \vdots & \vdots & & \vdots \\ y'_{m-(L+1)d}(2) & y'_{m-Ld}(2) & y'_{m-(L-1)d}(2) & \cdots & y'_{m-d}(2) \\ \hline & & \vdots & & \\ \hline y'(N) & y'_d(N) & y'_{2d}(N) & \cdots & y'_{Ld}(N) \\ y'_d(N) & y'_{2d}(N) & y'_{3d}(N) & \cdots & y'_{(L+1)d}(N) \\ \vdots & \vdots & \vdots & & \vdots \\ y'_{m-(L+1)d}(N) & y'_{m-Ld}(N) & y'_{m-(L-1)d}(N) & \cdots & y'_{m-d}(N) \end{array}\right]. \tag{6}$$

Equation (5) is the set of backward linear prediction equations formed from the decimated subset of data. Here L is the order of prediction, and b is the coefficient vector of the polynomial $B(z)$ given by

$$B(z) = 1 + b_1 z + b_2 z^2 + \cdots + b_L z^L. \tag{7}$$

Ideally, 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 $\begin{bmatrix} y' & Y' \end{bmatrix}$ and truncating all but the first n singular values to arrive at a noise cleaned estimate $\begin{bmatrix} \widehat{y'} & \widehat{Y'} \end{bmatrix}$.[1] This leads to the modified linear prediction equation

$$\widehat{Y'}\widehat{b} = -\widehat{y'} \tag{8}$$

from which the linear prediction coefficient vector estimate $\widehat{b}$ is found as $\widehat{b} = -\widehat{Y'}^{+}\widehat{y'}$, where $^+$ denotes the Moore-Penrose pseudoinverse. Finally, the estimates for the decimated poles are found by

$$\widehat{p_j^d} = \text{zero}_j\left(\widehat{B}(z)\right), \qquad j = 1, 2, \ldots, L. \tag{9}$$

The estimates for the undecimated poles are then given by

$$\widehat{p}_j = \widehat{p_j^d}^{\frac{1}{d}}. \tag{10}$$

Once the poles have been determined, the amplitude coefficients are estimated. Equation (3) leads to the following equation for the amplitude coefficients,

$$\begin{bmatrix} 1 & 1 & \cdots & 1 \\ \widehat{p}_1 & \widehat{p}_2 & \cdots & \widehat{p}_L \\ \widehat{p}_1^2 & \widehat{p}_2^2 & \cdots & \widehat{p}_L^2 \\ \vdots & \vdots & & \vdots \\ \widehat{p}_1^{(m-1)} & \widehat{p}_2^{(m-1)} & \cdots & \widehat{p}_L^{(m-1)} \end{bmatrix} \begin{bmatrix} \widehat{x}_1(1) & \widehat{x}_1(2) & \cdots & \widehat{x}_1(N) \\ \widehat{x}_2(1) & \widehat{x}_2(2) & \cdots & \widehat{x}_2(N) \\ \vdots & \vdots & & \vdots \\ \widehat{x}_L(1) & \widehat{x}_L(2) & \cdots & \widehat{x}_L(N) \end{bmatrix} = \begin{bmatrix} y'(1) & y'(2) & \cdots & y'(N) \end{bmatrix} \tag{11}$$

or

$$\widehat{A}_L \widehat{X} = Y'_a. \tag{12}$$

The amplitude coefficients can be found from a least squares solution to Equation (12),

$$\widehat{X} = \left(\widehat{A}_L^* \widehat{A}_L\right)^{-1} \widehat{A}_L^* Y_a', \tag{13}$$

where * denotes complex conjugate transpose. Because only n' singular values of $\widehat{Y}$ 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. We then reestimate the amplitude coefficients of these n poles. This is done using Equation (13), 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 Equation (13) is not used in practice to solve Equation (11), as more numerically sound procedures (such as QR decomposition[10]) can be used.

3. STATISTICAL ANALYSIS

A major contribution of this paper is a theoretical expression for the statistical properties of the estimated poles ($\widehat{p}_i$). Below we derive a general expression for the asymptotic pdf of these estimates. This expression applies to different decimation values so it can be used to determine the relative statistical accuracy for various choices in the TLS-Prony algorithm.

To analyze pole statistics we make the assumption that an ideal filter is applied to the data which completely eliminates modes outside the band of interest (*i.e.* those with frequencies outside the low-pass band $f \in \left[-\frac{1}{2d}, \frac{1}{2d}\right]$; this estimation can be generalized to provide estimates from any contiguous band, as shown in Section 4). This ideal filter also reduces the noise power by a factor of d. Note that the filtered noise in the decimated sequence remains white, since the decimation operation spreads out the low-pass noise to cover the full unambiguous frequency spectrum (see Section 4).

To analyze pole statistics for the poles within the band of interest we now derive their asymptotic pdf. This is given in the following theorem.

Theorem: Assume we are given data $y(t)$ as defined in Equations (1) and (2). Let

$$P_d = \left[\begin{array}{cccc} p_1^d & p_2^d & \cdots & p_{n'}^d \end{array}\right]^T \tag{14}$$

be the n' model poles which lie in the frequency band $f \in \left[-\frac{1}{2d}, \frac{1}{2d}\right]$, and let $\widehat{P}_d$ denote the TLS-Prony estimate of P_d using decimation, which is given by the n' highest energy pole estimates found in Equation (9). Then the asymptotic (high SNR, where SNR is defined as total signal power divided by total noise power) pdf of $\widehat{P}_d$ is given by

$$\widehat{P}_d \sim N\left(P_d, \Sigma_d\right), \tag{15}$$

where

$$\Sigma_d \triangleq \mathrm{Cov}\left(\widehat{P}_d\right) = \frac{\sigma}{d} F_d G_d S'^{+} B_d B_d^* \left(F_d G_d S'^{+}\right)^*. \tag{16}$$

Here

$$B_d = \left[\begin{array}{ccccccccc} 1 & b_1 & b_2 & \cdots & b_L & 0 & 0 & \cdots & 0 \\ 0 & 1 & b_1 & \cdots & b_{L-1} & b_L & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 1 & b_1 & \cdots & b_{L-1} & b_L & 0 \\ 0 & \cdots & 0 & 0 & 1 & \cdots & b_{L-2} & b_{L-1} & b_L \end{array}\right]_{(m_d - L)\times(m_d)}, \tag{17}$$

where $m_d = \frac{m}{d}$,

$$F_d = \mathrm{diag}\left(\frac{1}{\gamma_1}, \frac{1}{\gamma_2}, \ldots, \frac{1}{\gamma_{n'}}\right), \tag{18}$$

$$\gamma_i = \begin{bmatrix} b_1 & b_2 & \cdots & b_L \end{bmatrix} \begin{bmatrix} 1 \\ 2p_i^d \\ \vdots \\ Lp_i^{d(L-1)} \end{bmatrix}, \tag{19}$$

$$G_d = \begin{bmatrix} p_1^d & p_1^{2d} & \cdots & p_1^{Ld} \\ p_2^d & p_2^{2d} & \cdots & p_2^{Ld} \\ \vdots & \vdots & & \vdots \\ p_{n'}^d & p_{n'}^{2d} & \cdots & p_{n'}^{Ld} \end{bmatrix}, \tag{20}$$

and S' is the noise free version of Y'. □

Proof: See Appendix.

Equation (16) provides the covariances for decimated pole estimates given a particular set of poles and decimation factor. The variances for the undecimated poles can be easily derived in terms of the variances for the decimated poles using a first order approximation. The relationship is found as

$$\mathrm{Var}(\widehat{p}) = \frac{\mathrm{Var}\left(\widehat{p^d}\right)}{d^2 \left|p\right|^{2(d-1)}}. \tag{21}$$

The variances of the undecimated poles can now be compared to their respective CRBs. Recently, a CRB formulation has been developed for multi-snapshot damped exponentials in noise.[11] These CRB results can be directly compared with the variances of the estimated poles using the TLS-Prony method to examine its performance in both undecimated and decimated circumstances. This comparison is shown in Section 5 for a number of examples.

4. FULL SPECTRUM ESTIMATION USING FILTERING AND DECIMATION

Using the decimation scheme which has been presented, any poles or modes not in the area of interest need to be filtered out so that they are not aliased into the band of interest by the decimation operation. Even if there are no poles outside the band of interest a filter is still applied to reduce the noise in these bands which would also be overlapped into the band of interest in the decimation operation. It is important to note that the statistical analysis above applies only to the case where an ideal filter is used (*i.e.* one which eliminates the modes outside the band of interest and ideally low-passes the noise). The discussion below shows the effects of filtering and decimation and develops a method for repeated use of the procedure to analyze full spectrum data sets.

To examine the use of filtering and decimation, assume we are interested in estimating complete spectrum using a decimation factor of d. We estimate poles in each region of Figure 1(a) using a decimation based TLS-Prony procedure. First, we modulate the data to center the band of interest about $f = 0$ as follows

$$y_q^{mod}(t) = y_q(t) e^{-j2\pi f_0 q}, \tag{22}$$

where $f_0 = \frac{k}{d}$ is the modulating frequency for the kth subband, $k = 0, 1, \ldots, d-1$. We then lowpass filter the modulated data, $y^{mod}(t)$, to isolate the frequency band $f \in \left[-\frac{1}{2d}, \frac{1}{2d}\right]$. Finally, we apply the decimated TLS-Prony algorithm of Section 2. The resulting pole estimates, $\widehat{p_j^d}$, as given by Equation (9) lie in $f \in \left[-\frac{1}{2}, \frac{1}{2}\right]$ as shown in Figure 1(b). The corresponding pole estimates in undecimated frequency space are given by Equation (10) with modulation as follows

$$\widehat{p}_j = \widehat{p_j^d}^{\frac{1}{d}} e^{j2\pi f_0}. \tag{23}$$

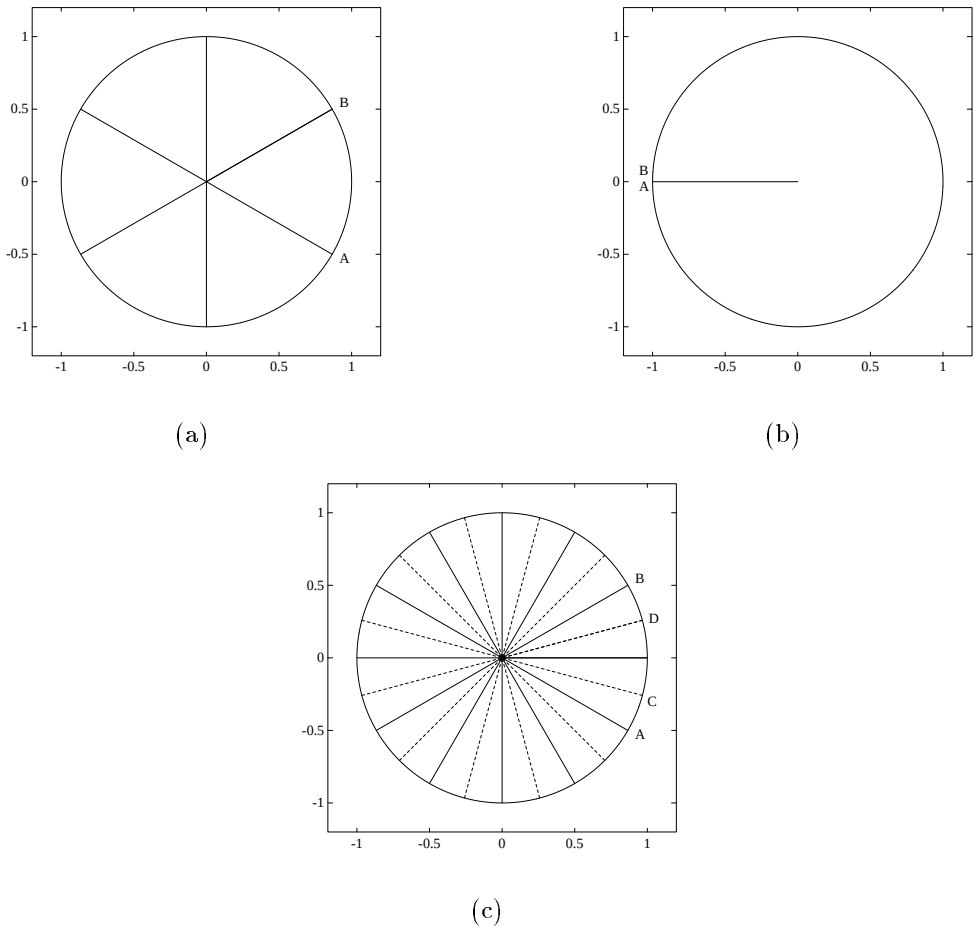


Figure 1: Spectral effect due to decimating by six.

A problem which can result from the above procedure is that poles near the end points of the subband region may be incorrectly estimated at near the opposite end point. This results from the discontinuity of the mapping in Equation (23) for $\angle \widehat{p_j^d} \approx \pm\pi$. It can be seen from Figure 1(b) that small errors in estimates near A and B result in large differences in Figure 1(a). To avoid this problem, we use $k > d$ overlapped estimates, each of size $\frac{1}{d}$, as shown in Figure 1(c) for $k = 2d$. We divide the overlap regions in half, and retain pole estimates from only one of the subbands in this region. For example, in Figure 1(c) for the subband [A,B] we retain poles only in the region [C,D]. This corresponds to retaining pole estimates whose angles satisfy

$$\left| \angle \widehat{p_j^d} \right| \in \left[0, \frac{\pi}{k} \right] \tag{24}$$

in the decimated frequency space. We note that this overlap method provides immunity to effects of a nonideal lowpass filter, as is seen in the examples in Section 5. In those examples we use an lth order equiripple FIR filter with a cutoff frequencies at the points C and D and stopband frequencies at the points A and B. After filtering, the l points at the beginning of the data are discarded to avoid filtering transients. Of course, the smaller k is, the smaller the required transition band will be and thus the longer the filter will need to be. We can see that there are design tradeoffs between pole estimation performance, filter order, and number of subestimations.

Once all pole estimates have been found, the amplitude coefficient estimates are given by Equation (13). Note that the amplitude coefficients are not found from the filtered data to avoid the effects of the lowpass filter operation on the pole amplitudes. As a result, only a single amplitude estimate computation is performed.

A comparison between the computations of the undecimated TLS-Prony and decimated TLS-Prony methods reveals that decimation can provide significant savings. An SVD operation count is presented here for the case when the data is real. For complex data considered in the examples which follow the counts were observed to be about a factor of 2 to 3 larger.

Assume the model order used for estimation is an integer near $L = \frac{m_d}{3} = \frac{m}{3d}$ (this choice is motivated by the fact that $L \approx \frac{m_d}{3}$ gives near optimal performance; see Example 1 below. This leads to a $\left[\begin{array}{cc} \widehat{y'} & \widehat{Y'} \end{array}\right]$ prediction matrix which has dimension $r \times c = \left(\frac{2}{3}\frac{mN}{d}\right) \times \left(\frac{m}{3d}+1\right)$. The approximate floating point operation (flop) count associated with the "economical" SVD computation (i.e. one in which only the first c left singular vectors are computed) is given by $\text{fc} \approx 14rc^2 + 8c^3$.[10]

For the undecimated case we thus get an SVD flop count of

$$\text{fc}_{\text{undec}} \approx 14\left(\frac{2}{3}mN\right)\left(\frac{m}{3}+1\right)^2 + 8\left(\frac{m}{3}+1\right)^3. \tag{25}$$

For the decimated cases, recall that the data length has been reduced by l and that k SVDs are required. We thus get

$$\text{fc}_{\text{dec}} \approx k\left(14\left(\frac{2}{3}\,\frac{m-l}{d}N\right)\left(\frac{m-l}{3d}+1\right)^2 + 8\left(\frac{m-l}{3d}+1\right)^3\right). \tag{26}$$

The computational savings for particular choices of d, k, and l can now be found as $\text{fc}_{\text{undec}}/\text{fc}_{\text{dec}}$ for most practical choices of the estimation parameters. Note that $\text{fc}_{\text{undec}}/\text{fc}_{\text{dec}}$ is significantly larger than 1 (see Example 3 below). It is also noted that in addition to computational savings, there are significant memory savings due to the smaller prediction matrices which must be decomposed.

5. SIMULATION STUDIES

Simulations have been performed which demonstrate the advantages of using decimation from results of statistical analyses. Simulations are also performed on full spectrum data sets to demonstrate the estimation ability of the modified TLS-Prony method which has been presented.

5.1 Example 1

In this simulation we assume one snapshot of data of length $m = 140$. We also assume a single exponential located on the unit circle and an SNR of 5dB. Figure 2 shows the theoretical variance of the estimated pole versus prediction order for various decimation factors as compared to the CRB.

From this figure we can see that the minimum variance occurs at a lower prediction order for higher d. The minimum occurs for a prediction order equal to about one third of the decimated data length (i.e. $L = \frac{m_d}{3}$), which is consistent with results for undecimated data.[12] This shows that we can obtain optimal performance for the decimated cases a lower prediction orders than for the undecimated case, thus reducing the required computation.

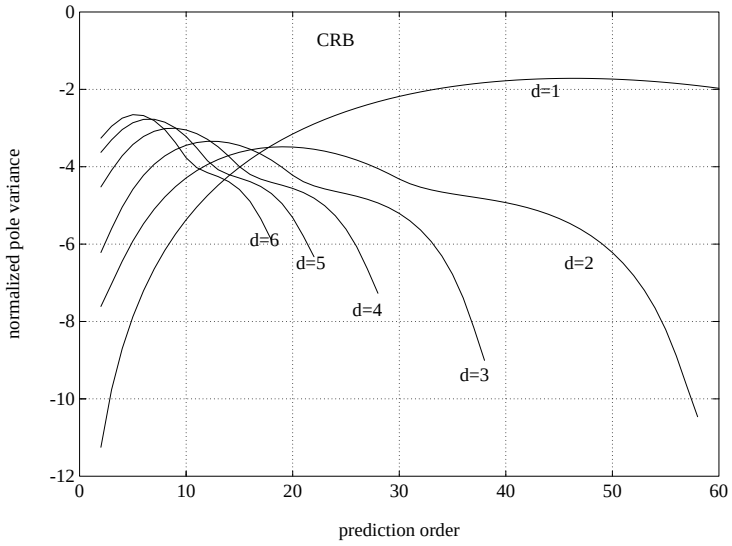


Figure 2: $10\log_{10}(\mathrm{CRB}/\mathrm{Var}(p))$ for various d and prediction orders for a single pole on the unit circle.

We note, however, that since the data has to be filtered prior to decimation, the curves for the decimated cases move down according to the number of data points which need to be discarded after filtering. In this example, for instance, if a 20 point FIR filter were used, there would be 120 data points remaining for decimated estimation and the decimated curves were observed to move down about 1.5dB. Thus we have a performance tradeoff for reduced computation. It is important to note that this shift becomes smaller as the data length is increased since the percent difference between the original and filtered data lengths decreases. We also point out that this implies that there is a tradeoff between the performance shift and filter length for the same reason: the shift becomes smaller as filter length is reduced. Note, however, that reducing the filter length reduces its rejection quality in practice since the filter is nonideal.

5.2 Example 2

In a second simulation we make the same assumptions above, except that there are two equal energy exponentials located on the unit circle one Fourier bin apart (i.e. $\Delta f = \frac{1}{m} = \frac{1}{140}$). The total SNR was 8dB in this case in order to maintain 5dB SNR/pole. Figure 3 shows the theoretical variance of the estimates for one of the poles (the variance of the other pole is identical). We can see that the characteristics are much the same as in the one pole case, the difference being higher variances due to the presence of each pole's neighbor.

It is thus clear from Examples 1 and 2 that we can reduce computational cost by using decimation with filtering, with a sacrifice of accuracy. The computational savings and performance of full spectrum estimation using filtering and decimation can be seen in Example 3 below.

5.3 Example 3

Another set of simulations have been performed on a full spectrum set of data. In these simulations we had $N = 1$ snapshot, and $n = 10$ poles were present in the data. The amplitude coefficients all have unit magnitude; the phases of the amplitude coefficients were chosen randomly. In these simulations $m = 140$ data points were used. Figure 4(a) shows the locations of the ten poles; each is indicated by an "x". Two-hundred independent Monte-Carlo

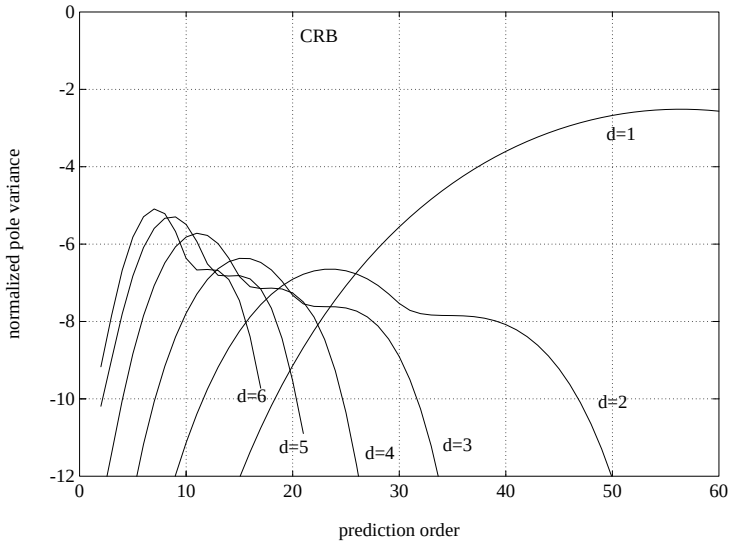
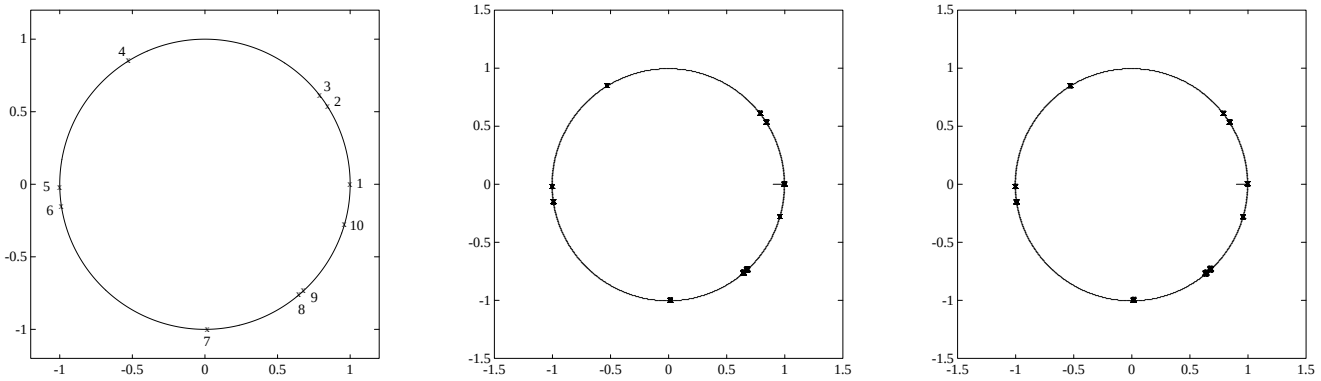


Figure 3: $10\log_{10}(\mathrm{CRB}/\mathrm{Var}(p_1))$ for various d and prediction orders for two poles on the unit circle.



(a) (b) (c)

Figure 4: Example 2: (a) true pole locations, (b) estimated pole locations for $d = 1$, and (c) estimated pole locations for $d = 6$.

simulations are performed by adding noise to the data such that the total SNR was 15dB, again in order to maintain 5dB/pole. Estimates for the poles are obtained using the TLS-Prony algorithm without decimation (*i.e.* $d = 1$) and with decimation using a decimation factor of $d = 6$.

For the decimated results, $k = 12$ overlapping subestimations were performed using an $l = 20$ order equiripple FIR filter. Only the last 120 data points of the filter output are used, as the first 20 points are corrupted by the transient response of the filter. The prediction order and number of singular values retained in the simulations were $L = 47$ and 10, for $d = 1$, and $L = 7$ and 2, respectively, for each of the twelve subestimations for $d = 6$. These prediction orders correspond to one third of the effective data lengths in the two cases as was suggested by Examples 1 and 2. Prior to the calculation of the amplitude coefficients, poles with magnitude larger than 1.3 and smaller than $\frac{1}{1.3}$ were eliminated to avoid poor conditioning in the least squares solution of the amplitude coefficients.

pole number	CRB	$d=1$			$d=6$		
		Theory	Simulation	Sim. MSE	Theory	Simulation	Sim. MSE
1	-58.4	-57.1	-56.8	-56.8	-55.1	-56.5	-49.9
2	-58.0	-56.0	-55.9	-55.9	-54.4	-54.7	-54.5
3	-58.0	-55.9	-56.1	-56.1	-54.4	-54.9	-54.4
4	-58.6	-56.6	-56.4	-56.3	-55.1	-54.0	-53.6
5	-58.4	-56.9	-57.1	-57.0	-54.8	-55.2	-51.1
6	-58.4	-57.0	-56.6	-56.5	-54.8	-55.2	-51.6
7	-58.5	-56.2	-55.4	-55.4	-55.1	-53.1	-51.8
8	-50.1	-46.8	-46.8	-46.8	-43.4	-45.1	-41.2
9	-50.1	-46.7	-46.5	-46.5	-43.4	-45.5	-43.0
10	-58.4	-56.9	-57.2	-57.2	-55.1	-54.8	-51.4

Table 1: Theoretical and simulation variances and MSEs for the poles in Example 2 (all values are in dBs).

The estimated pole location results for $d = 1$ and $d = 6$ are shown in Figures 4(b) and (c), respectively. The "x" symbols indicate the ten highest energy pole estimates for each Monte-Carlo simulation. From these figures we can see that the poles are well estimated for both the $d = 1$ case and the $d = 6$ case. The details of the pole estimation accuracy are summarized in Table 1. The relative variances (relative to the CRB) for the poles in the simulations for the $d = 1$ case vary from -1.2dB to -3.6dB. Note that simulation variances are also very close to those predicted by the theory (the differences are due to nonideal filtering). Note also that the estimates have negligible bias which is shown by the fact that the pole variances are very close to their MSEs (MSE is defined as the variance plus the square of the bias).

The relative variances for the $d = 6$ case vary from -1.9dB to -5.4dB, which are slightly worse than those for the undecimated case, as predicted by Example 1. Looking at the MSEs, we can see that the decimated estimates are also biased. We can thus see the performance sacrifice made for computation reduction detailed next.

These two estimation procedures are now compared on the basis of their computational costs. Using MATLAB, the "economical" version of the SVD operation for the undecimated data matrix required an average of 5.1Mflops for each of the Monte-Carlo simulations. Each of the twelve $d = 6$ SVDs required an average of 31Kflops, resulting in a total of 372Kflops for each Monte-Carlo simulation. The computational savings for the SVDs in this example using decimation was thus a factor of 13.7. This compares with a savings factor of 20.6 which is predicted by Equations (25) and (26) for this scenario. Note that these numbers cannot be computed directly in part because Equations (25) and (26) are approximations for the real case and perhaps in part due to differences between MATLAB's SVD algorithm and the one described by Golub and Van Loan.[10] With higher decimation factors the savings are even more substantial.

6. CONCLUSIONS

This paper has presented a new TLS-Prony estimation algorithm which incorporates data decimation. Also presented is a statistical analysis for estimated poles of this algorithm. We have shown through this analysis that decimation provides a minimum variance for estimated poles that occurs at a prediction order which is smaller than the optimal prediction order for undecimated data by a factor of d, thus allowing for computational savings. It is shown that this benefit is obtained at the expense of pole variance performance due to the filtering which is required; this expense becomes smaller for longer data lengths. We have also shown how the modified TLS-Prony method can be used on full spectrum data one band at a time to realize the computational savings in a general signal framework.

7. ACKNOWLEDGMENTS

This research was supported in part by the Air Force Office of Scientific Research, by the Avionics Division, Wright Laboratories, and by the Surveillance Division, Rome Laboratories. We would also like to thank William Pierson for his help with the simulations.

8. APPENDIX: PROOF OF THEOREM

From Equation (8) we can make the following substitutions

$$\left(S' + \tilde{S}'\right)\left(b + \tilde{b}\right) = -\left(s' + \tilde{s}'\right), \tag{27}$$

where s' is the noise free version of y', $\widehat{Y'} = S' + \tilde{S}'$, $\widehat{y'} = s' + \tilde{s}'$, and $\widehat{b} = b + \tilde{b}$. We can see that the $\tilde{\cdot}$ terms are small perturbations for the high SNR case, which is assumed. Multiplying out Equation (27) and retaining only the first order $\tilde{\cdot}$ terms gives approximately,

$$S'b + \tilde{S}'b + S'\tilde{b} = -s' - \tilde{s}'. \tag{28}$$

Now note that $S'b = -s'$ since they are the noiseless terms. Equation (28) thus becomes

$$S'\tilde{b} = -\left(\tilde{s}' + \tilde{S}'b\right). \tag{29}$$

Multiplying both sides through by $S'S'^{+}$ and noting that $S'S'^{+}S' = S'$, we obtain

$$S'\tilde{b} = -S'S'^{+}\left(\tilde{s}' + \tilde{S}'b\right). \tag{30}$$

Let $Y' = S' + W'$ and $y' = s' + w'$, where W' and w' are the appropriate noise matrices (i.e. of the form given in Equation (5)). Thus we can see that $\tilde{S}'$ and $\tilde{s}'$ are perturbations caused by W' and w' and the SVD truncation which is performed. By using perturbation analysis[13] on the matrices $\left[\begin{array}{cc} s' & S' \end{array}\right]$, $\left[\begin{array}{cc} y' & Y' \end{array}\right]$, and $\left[\begin{array}{cc} \widehat{y'} & \widehat{Y'} \end{array}\right]$, it can be shown that to first order approximation

$$S'^{+}\left[\begin{array}{cc} \tilde{s}' & \tilde{S}' \end{array}\right] = S'^{+}\left[\begin{array}{cc} w' & W' \end{array}\right]. \tag{31}$$

Thus, Equation (30) can be written as

$$S'\tilde{b} = -S'S'^{+}\epsilon, \tag{32}$$

where $\epsilon = w + Wb$.

Observing the data model and the formulation of the S' matrix, we can write S'^{l} as

$$S' = \left[\begin{array}{ccc|ccc|c|ccc} x_1(1) & x_2(1) & \cdots & x_{n'}(1) & & & & & & \\ x_1(1)p_1^d & x_2(1)p_2^d & \cdots & x_{n'}(1)p_{n'}^d & & & & & & \\ \vdots & \vdots & & \vdots & & & & & & \\ x_1(1)p_1^{m-(L+1)d} & x_2(1)p_2^{m-(L+1)d} & \cdots & x_{n'}(1)p_{n'}^{m-(L+1)d} & & & & & & \\ \hline x_1(2) & x_2(2) & \cdots & x_{n'}(2) & & & & & & \\ x_1(2)p_1^d & x_2(2)p_2^d & \cdots & x_{n'}(2)p_{n'}^d & & & & & & \\ \vdots & \vdots & & \vdots & & & & & & \\ x_1(2)p_1^{m-(L+1)d} & x_2(2)p_2^{m-(L+1)d} & \cdots & x_{n'}(2)p_{n'}^{m-(L+1)d} & & & & & & \\ \hline & & & & & & \vdots & & & \\ \hline x_1(N) & x_2(N) & \cdots & x_{n'}(N) & & & & & & \\ x_1(N)p_1^d & x_2(N)p_2^d & \cdots & x_{n'}(N)p_{n'}^d & & & & & & \\ \vdots & \vdots & & \vdots & & & & & & \\ x_1(N)p_1^{m-(L+1)d} & x_2(N)p_2^{m-(L+1)d} & \cdots & x_{n'}(N)p_{n'}^{m-(L+1)d} & & & & & & \end{array}\right] G_{\mathrm{d}}, \tag{33}$$

or

$$S' = HG_d. \tag{34}$$

Equation (32) thus becomes

$$HG_d\tilde{b} = -HG_dS'^{+}\epsilon. \tag{35}$$

Now note from Equation (7) that the true and estimated Lth order characteristic polynomials are $B(z) = 1 + b_1z^d + b_2z^{2d} + \cdots + b_Lz^{Ld}$ and $\widehat{B}(z) = 1 + \widehat{b}_1z^d + \widehat{b}_2z^{2d} + \cdots + \widehat{b}_Lz^{Ld}$, respectively. Note that $B\left(p_i^d\right) = 0$ and $\widehat{B}\left(\widehat{p_i^d}\right) = 0$.

We can use a first order Taylor expansion to find an expression for the error in the estimated pole locations. For each $\widehat{p_i^d}$ we obtain

$$\begin{aligned}
0 &= \widehat{B}\left(\widehat{p_i^d}\right) \\
&= \widehat{B}\left(p_i^d\right) + \frac{\partial}{\partial z}\widehat{B}(z)|_{z=p_i^d}\left(\widehat{p_i^d} - p_i^d\right) + (\text{higher order terms}) \\
&= \widehat{B}\left(p_i^d\right) - B\left(p_i^d\right) + \frac{\partial}{\partial z}\widehat{B}(z)|_{z=p_i^d}\left(\widehat{p_i^d} - p_i^d\right) + (\text{higher order terms}) \\
&\approx 1 + \widehat{b}_1p_i^d + \widehat{b}_2p_i^{2d} + \cdots + \widehat{b}_Lp_i^{Ld} - \left(1 + b_1p_i^d + b_2p_i^{2d} + \cdots + b_Lp_i^{Ld}\right) \\
&\quad + \left(\widehat{b}_1 + 2\widehat{b}_2p_i^d + \cdots + L\widehat{b}_Lp_i^{(L-1)d}\right)\left(\widehat{p_i^d} - p_i^d\right) \\
&\approx \begin{bmatrix} p_i^d & p_i^{2d} & \cdots & p_i^{Ld} \end{bmatrix}\begin{bmatrix} \widehat{b}_1 - b_1 \\ \widehat{b}_2 - b_2 \\ \vdots \\ \widehat{b}_L - b_L \end{bmatrix} + \begin{bmatrix} b_1 & b_2 & \cdots & b_L \end{bmatrix}\begin{bmatrix} 1 \\ 2p_i^d \\ \vdots \\ Lp_i^{(L-1)d} \end{bmatrix}\left(\widehat{p_i^d} - p_i^d\right) \\
&\approx \begin{bmatrix} p_i^d & p_i^{2d} & \cdots & p_i^{Ld} \end{bmatrix}\left(\widehat{b} - b\right) + \gamma_i\left(\widehat{p_i^d} - p_i^d\right),
\end{aligned} \tag{36}$$

or, approximately,

$$\left(\widehat{p_i^d} - p_i^d\right) = -\frac{1}{\gamma_i}\begin{bmatrix} p_i^d & p_i^{2d} & \cdots & p_i^{Ld} \end{bmatrix}\tilde{b}. \tag{37}$$

In matrix form for the n' true poles we thus obtain

$$\widehat{P}_d - P_d = -F_dG_d\tilde{b}. \tag{38}$$

Since H is full rank, we can multiply Equation (35) by $\left(H^*H\right)^{-1}H^*$ to get

$$G_d\tilde{b} = -G_dS'^{+}\epsilon. \tag{39}$$

and by substituting Equation (39) into Equation (38) we obtain

$$\widehat{P}_d - P_d = F_dG_dS'^{+}\epsilon. \tag{40}$$

Recall W' and w' are zero mean Gaussian and thus ϵ is multivariate Gaussian with zero mean and covariance matrix

$$\begin{aligned}
\text{Cov}(\epsilon) &= E\left[\left(\begin{bmatrix} w' & W' \end{bmatrix}\begin{bmatrix} 1 \\ b \end{bmatrix}\right)\left(\begin{bmatrix} w' & W' \end{bmatrix}\begin{bmatrix} 1 \\ b \end{bmatrix}\right)^*\right] \\
&= \frac{\sigma}{d}B_dB_d^*.
\end{aligned} \tag{41}$$

Since ϵ is zero mean, Equation (40) implies that $E\left[\widehat{P}_d\right] = P_d$. Equations (40) and (41) imply that the covariance matrix of $\widehat{P_d}$ is given by Equation (16). □

9. 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, no. 10, pp. 1440–1454, Oct. 1987.

2. W. M. Steedly and R. L. Moses, "High resolution exponential modeling of fully polarized radar returns," *IEEE Transactions on Aerospace and Electronic Systems*, vol. AES-27, no. 3, pp. 459–469, May 1991.

3. B. Liu and F. Mintzer, "Calcuation of narrow-band spectra by direct decimation," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. ASSP-26, no. 6, pp. 529–534, Dec. 1978.

4. M. P. Quirk and B. Liu, "Improving resolution for autoregressive spectral estimation by decimation," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. ASSP-31, no. 3, pp. 630–637, June 1983.

5. M. J. Villalba and B. K. Walker, "Spectrum manipulation for improved resolution," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. ASSP-37, no. 6, pp. 820–831, June 1989.

6. C.-H. J. Ying, R. L. Moses, and R. L. Dilsavor, "Perturbation analysis for pole estimates of damped exponential signals," tech. rep., The Ohio State University, Department of Electrical Engineering, ElectroScience Laboratory, Aug. 1991.

7. G. Xu, S. D. Silverstein, R. H. Roy, and T. Kailath, "Parallel implementation and performance analysis of beamspace ESPRIT," in *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, (Toronto, Ontario), pp. 1497–1500, May 14–17, 1991.

8. M. D. Zoltowski, G. M. Kautz, and S. D. Silverstein, "Development, performance analysis, and experimental evaluation of beamspace root-MUSIC," in *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, (Toronto, Ontario), pp. 3049–3052, May 14–17, 1991.

9. S. D. Silverstein, W. E. Engeler, and J. A. Taridif, "Parallel architectures for multirate superresolution spectrum analyzers," *IEEE Transactions on Circuits and Systems*, vol. 38, no. 4, pp. 449–453, Apr. 1991.

10. G. H. Golub and C. F. VanLoan, *Matrix Computations.* Baltimore, MD: Johns Hopkins, 2 ed., 1989.

11. 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.

12. 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, no. 5, pp. 814–824, May 1990.

13. Y. Hua and T. K. Sarkar, "A perturbation property of the TLS-LP method," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. ASSP-38, no. 11, pp. 2004–2005, Nov. 1990.