Estimating the Parameters of Exponentially Damped Sinusoids and Pole-Zero Modeling in Noise

RAMDAS KUMARESAN AND DONALD W. TUFTS, FELLOW, IEEE

Abstract—**We have presented techniques [1]-[6] based on linear prediction (LP) and singular value decomposition (SVD) for accurate estimation of closely spaced frequencies of sinusoidal signals in noise. In this note we extend these techniques to estimate the parameters of exponentially damped sinusoidal signals in noise. The estimation procedure presented here makes use of "backward prediction" in addition to SVD.**

First, the method is applied to data consisting of one and two exponentially damped sinusoids. The choice of one and two signal components facilitates the comparison of estimation error in pole damping factors and pole frequencies to the appropriate Cramer-Rao (CR) bounds and to traditional methods of linear prediction. Second, our method is applied to an example due to Steiglitz [8] in which the data consists of noisy values of the impulse response samples (composed of many exponentially damped sinusoids) of a linear system having both poles and zeros. The poles of the system are accurately determined by our method and the zeros are obtained subsequently, using Shanks' method.

I. Introduction

We have presented techniques [1]-[6] based on linear prediction and singular value decomposition for accurate estimation of closely spaced sinusoidal signals in noise. These techniques improved the least-squares linear prediction methods proposed by Ulrych and Clayton [9] and Nuttall [10] which in turn were improvements of Prony's original method [11]. We were able to improve the performance of these techniques considerably with two interrelated modifications. First, we applied singular value decomposition (SVD) to the linear prediction equations (2) and (3) and (5)-(7) to alleviate severe ill-conditioning. Second, we used large values of the prediction filter order L (a large fraction of N, the number of data samples). The number of sinusoidal signals M, whose frequencies are to be estimated was small compared to L. In this paper our purpose is to apply these techniques to estimate the parameters of exponentially damped sinusoidal signals in noise. Henderson's paper [12] is related to the work reported here.

In the next section we outline theoretical results which provide a rationale for our method, which includes the use of backward prediction. In Section III we compare the results obtained by our method to traditional methods using noisy exponentially damped sinusoids which represent pitch synchronous simulated voiced speech. The sample variance of the parameter estimates are also compared to the Cramer-Rao (CR) bounds for the pole damping factors and the pole frequencies. We also apply our method to the noisy values of the impulse response of a linear system with both poles and zeros. The poles are accurately found by our method and the zeros of the system are then obtained by using Shanks' method [13].

Manuscript received October 30, 1981; revised March 8, 1982, and June 3, 1982.

R. Kumaresan was with the Department of Electrical Engineering, University of Rhode Island, Kingston, RI 02881. He is now with GTE Business Communication Systems, McLean, VA.

D. W. Tufts is with the Department of Electrical Engineering, University of Rhode Island, RI 02881.

II. Main Results

Suppose that the N samples of the observed data sequence $y(n)$ consists of samples of M exponentially damped signals in complex valued white Gaussian noise $w(n)$.

$$y(n) = \sum_{k=1}^{M} a_k e^{s_k n} + w(n) \qquad n = 0, 1, \cdots, N-1 \tag{1}$$

where $s = -\alpha_k + j2\pi f_k$, $k = 1, 2, \cdots, M$ are complex numbers (α_k is positive) and a_k, $k = 1, 2, \cdots, M$ are the complex amplitudes. α_k's are the pole damping factors and f_k's are the pole frequencies. We shall set up the following linear prediction equations using complex conjugate data in the backward direction.

$$\begin{bmatrix} y^*(1) & y^*(2) \cdots y^*(L) \\ y^*(2) & y^*(3) \cdots y^*(L+1) \\ \vdots & \vdots \quad\quad \vdots \\ y^*(N-L) & y^*(N-L+1) \cdots y^*(N-1) \end{bmatrix} \begin{bmatrix} b(1) \\ b(2) \\ \vdots \\ b(L) \end{bmatrix} = - \begin{bmatrix} y^*(0) \\ y^*(1) \\ \vdots \\ y^*(N-L-1) \end{bmatrix} \tag{2}$$

$$Ab = -h$$

where b is the vector of backward prediction coefficients. "$*$" denotes complex conjugate. We can write the above equation in augmented form as $A'b' = 0$ where $A' = (h \mid A)$ and $b' = (1, b^T)^T$. "T" denotes matrix transpose. b' is the vector of prediction error filter coefficients. If the data is noiseless, since b' lies in the null space of A', it is easy to show that [14] the prediction-error filter polynomial

$$B(z) = 1 + b(1)\, z^{-1} + b(2)\, z^{-2} + \cdots + b(L)\, z^{-L} \tag{3}$$

will have zeros at $e^{-s_k^*}$, $k = 1, 2, \cdots, M$ if L is chosen to satisfy the inequality $M \leqslant L \leqslant N - M$. From these zeros one can

0096-3518/82/1200-0833$00.75

find the e^{s_k}'s by reflecting them inside the unit circle. These zeros, called signal zeros, fall outside the unit circle because of the way (2) is written using the data in reversed time direction. If $L > M$, clearly, $B(z)$ has $L - M$ other zeros called extraneous zeros. Also, if $L > M$, (2) has more than one solution (assuming that the data are noiseless) because the rank of A (or A') is M ($<L$). However, if we find the unique minimum norm solution to (2) (which minimizes $\|b\|^2 = |b(1)|^2 + |b(2)|^2 + \cdots + |b(L)|^2$) then it can be shown [14] that the $L - M$ extraneous zeros will always fall with in the unit circle. The significance of this fact should become clearer in the sequel. An outline of the proof is as follows. We can factor $B(z)$ as $B(z) = B_1(z) B_2(z)$ where $B(z)$ is an Mth degree polynomial with M signal zeros. Such a factor does exist because $B(z)$ has zeros at $e^{-s_k^*}$, $k = 1, 2, \cdots, M$. $B_2(z)$ has the $L - M$ extraneous zeros, in whose locations we are interested. Minimizing the norm of b in (2) is same as minimizing $\int_{-\pi}^{\pi} |B(e^{j\omega})|^2 \, d\omega$ since the first coefficient of $B(z)$ is unity. Since $B_1(z)$ is fixed, minimizing $\int_{-\pi}^{\pi} |B(e^{j\omega})|^2 \, d\omega$ is equivalent to solving the standard "linear prediction" problem (autocorrelation method [15]) where $B_2(z)$ is the "prediction error filter" and $B_1(z)$ corresponds to the Z transform of the "data sequence" and $B(z)$ is the Z transform of the "error sequence." See [14] for more details. We then know from well known results [15] that $B_2(z)$, the "prediction error filter" is minimum phase. In summary, the M signal zeros of $B(z)$ fall outside the unit circle at $e^{-s_k^*}$, $k = 1, 2, \cdots, M$ whereas the $L - M$ extraneous zeros fall inside the unit circle. This fact facilitates the identification of the M signal zeros from the $L - M$ extraneous zeros of $B(z)$. This is the primary reason for using the data in the backward direction in (2). The reason for using a polynomial of degree $L > M$ in the first place, is to increase the accuracy of the pole location estimates, i.e., estimates of e^{s_k}'s, as we shall see in the next section. Although the above results are only true for the ideal situation of noiseless data, as we shall show experimentally, they turn out to be valid for moderately noisy data as well. But the coefficients of $B(z)$ have to be accurately determined using the truncated SVD [16] as explained below.

In the presence of noise in the data if we attempt to solve (2) in the least square sense as is done in covariance method [15] there are considerable perturbations introduced in the b vector. The reason for this is that for $L > M$, $L - M$ columns of A tend to be closely dependent, causing a very unstable least square problem. See [5]-[7] for more explanations. But moderately large values of L are essential in improving the accuracy of the pole location estimates as we shall see in the next section. Therefore, to alleviate this ill-conditioning problem we make use of the SVD of A. We compute the SVD of A, but instead of finding the least square solution, we find a truncated SVD solution [16] by setting the smaller singular values of A to zero. That is we compute b as follows:

$$b = -\sum_{k=1}^{M} \sigma_k^{-1} [u_k^{\dagger} h] \, v_k \tag{4}$$

where σ_k, $k = 1, 2, \cdots, L$ or $N - L$ (depending on the rank of A) are the singular values of A. v_k, $k = 1, 2, \cdots, L$ and u_k, $k = 1, 2, \cdots, N - L$ are the eigenvectors of $A^{\dagger}A$ and $AA^{\dagger}$, respectively. "$\dagger$" stands for matrix complex conjugate transpose. The effect of using a truncated SVD is to increases the SNR in the data prior to obtaining the solution vector b. See [2], [4] for more details. In the above we have assumed that the value of M is known. Otherwise, it can be estimated from the size of the singular values of A. In the case of noiseless data, since the rank of A is only M, only $\sigma_1, \sigma_2, \cdots, \sigma_M$ will be nonzero and b in (4) will be the minimum norm solution as desired. In (4), we have made use of only the M principal eigenvectors of $A^{\dagger}A$ and $AA^{\dagger}$ which are more robust to the noise perturbations in the data. The result is that the $L - M$ extraneous zeros of $B(z)$ also tend to be less perturbed. Note that the least squares solution to (2) can be written in the same way as in (4) but it would include all the nonzero singular values of A and the corresponding eigenvectors of $A^{\dagger}A$ and $AA^{\dagger}$. Thus the stability in the coefficient vector b is achieved by dropping from the least square solution to (2) the less robust eigenvectors of $A^{\dagger}A$ and $AA^{\dagger}$.

III. Experimental Results

Three simulation experiments are used in this section to point out the superiority of the SVD method described above and the advantage of using backward prediction.

Experiment 1

The simulated data are given by the formula

$$y(n) = a_1 e^{s_1 n} + a_2 e^{s_2 n} + w(n) \qquad n = 0, 1, \cdots, 24 \tag{5}$$

where $s_1 = -\alpha_1 + j2\pi f_1 = -0.1 + j2\pi(0.52)$, $s_2 = -\alpha_2 + j2\pi f_2 = -0.2 + j2\pi(0.42)$, $a_1 = 1$, $a_2 = 1$. $M = 2$. The sequence $w(n)$ is white, and complex Gaussian, with variance $2\sigma^2$. SNR is $10 \log (1/2\sigma^2)$. SNR = 20 dB. Note that this is peak SNR. Fig. 1(a)-(i) show the zeros of the polynomial $B(z)$ calculated by different methods. Forty independent trials using different $w(n)$ sequences are performed in each case. The zeros of $B(z)$ obtained in each trial are superposed with respect to the unit circle. First, we applied the traditional Prony method [11], where $B(z)$ had two ($L = M = 2$) unknown coefficients $b(1)$ and $b(2)$. They were found by minimizing the error

$$E = \sum_{n=2}^{24} \left| y(n) - \sum_{k=1}^{2} b(k) \, y(n-k) \right|^2.$$

Fig. 1(a) shows the zeros of $B(z)$ for the forty trials. The true pole locations e^{s_1} and e^{s_2} are shown by cross marks. Fig. 1(b) and 1(c) show the results obtained by the standard covariance [15]. The polynomial $B(z)$ is allowed to have $L = 4$ and $L = 8$ unknown coefficients. That is the error

$$E = \sum_{n=L}^{24} \left| y(n) - \sum_{k=1}^{L} b(k) \, y(n-k) \right|^2$$

is minimized by choosing $b(k)$'s. Clearly the accuracy of the zero clusters around the true pole locations improves with increase in L. But the polynomial $B(z)$ has $L - M$ extraneous zeros (2 for $L = 4$ and 6 for $L = 8$) which can not be identified from the two signal zeros without prior information or the use of many 25 sample blocks of data. As L is increased, although the accuracy of the signal zeros clusters increases, the ill-

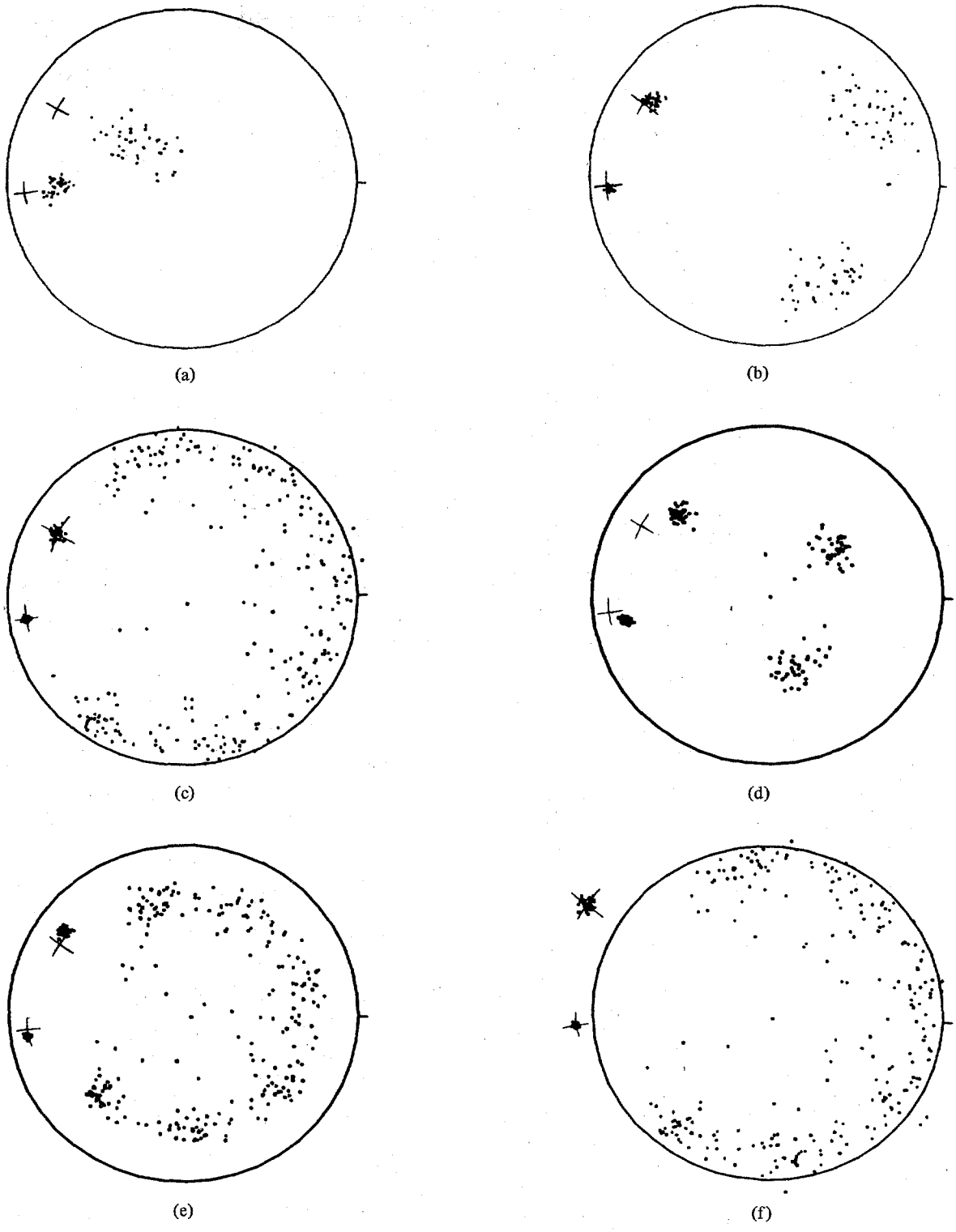

Fig. 1. The zeros of $B(z)$ obtained in forty independent trials are superposed with respect to the unit circle. The intersecting radial lines and arcs show the true locations of the two zeros. If the intersecting lines are inside the unit circle they correspond to e^{s_1} and e^{s_2} and if they are outside they correspond to $e^{-s_1^*}$ and $e^{-s_2^*}$. (a) Prony method, $L = M = 2$. (b) Covariance method, $L = 4$. (c) Covariance method, $L = 8$. (d) Autocorrelation method, $L = 4$. (e) Autocorrelation method, $L = 8$. (f) Backward covariance method, $L = 8$.

conditioning in the normal equations shows up as wild fluctuations in the locations of extraneous zeros. Fig. 1(d)-(e) show the results obtained by using the autocorrelation method of linear prediction [15]. This method is computationally simple and all the zeros are guaranteed to be inside the unit circle. But the accuracy in the pole estimates is very poor primarily due to the large bias. This method will not give correct results even in the absence of noise for short data records.

Fig. 1(f) shows the results when the backward covariance method is used. That is, (2) is solved in the least square sense

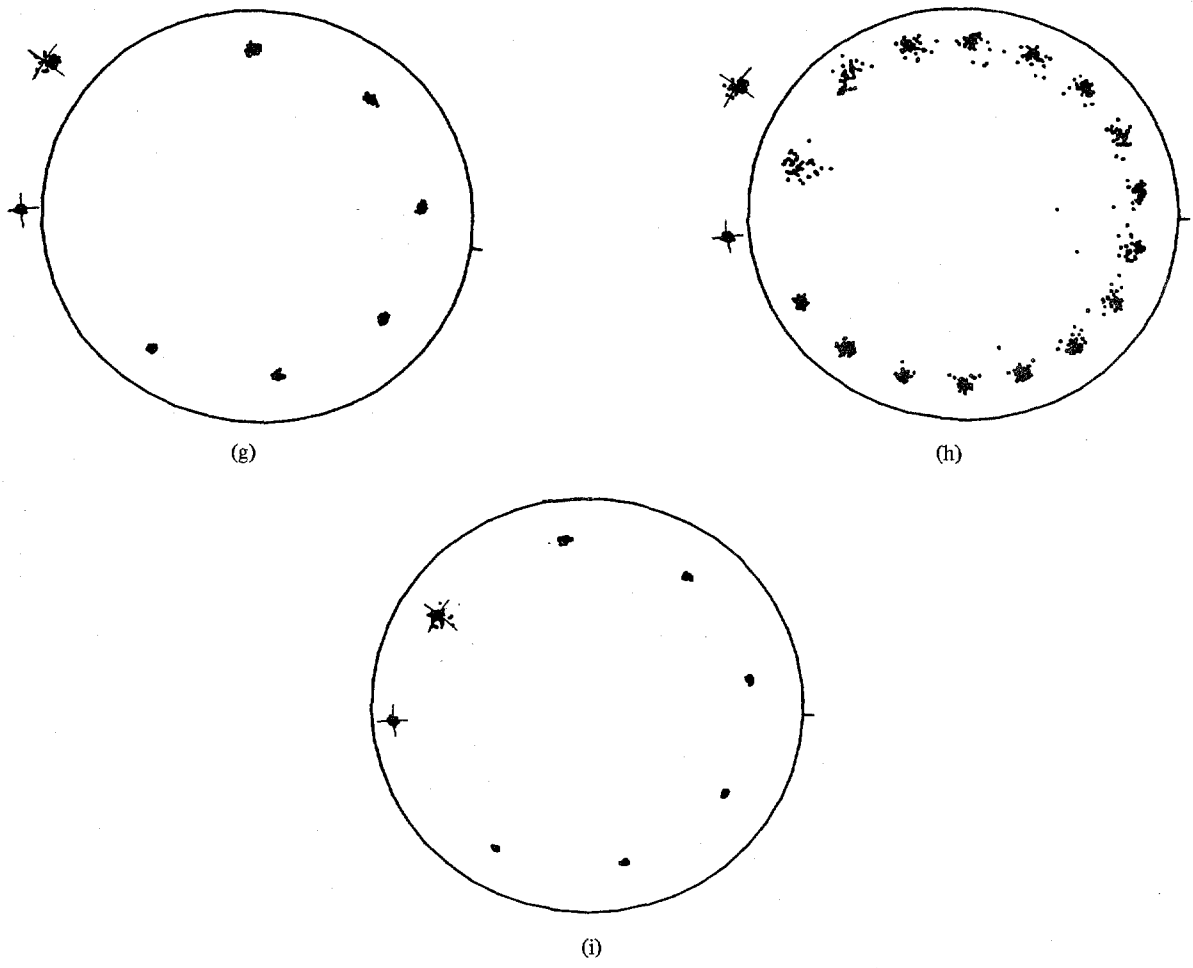

Fig. 1. (*Continued.*) (g) Backward covariance method using SVD, $L = 8$. (h) Backward covariance method using SVD, $L = 18$. (i) Covariance method using SVD, $L = 8$.

to find the coefficients of $B(z)$. The error minimized is

$$E = \sum_{n=0}^{24-L} \left| y^*(n) - \sum_{k=1}^{L} b(k)\, y^*(n+k) \right|^2.$$

Clearly, the signal zeros fall close to $e^{-s_k^*}$, $k = 1, 2$, but extraneous zeros occasionally fall outside the unit circle. Fig. 1(g)-(h) show the case when our SVD method is used to compute the backward-prediction coefficients of $B(z)$ for two different values of L. Clearly the reduction in the ill-conditioning or the improvement in the SNR of the data matrix, has resulted in better approximation to the noiseless case, and the extraneous zeros are much less perturbed from their noiseless locations [14]. Therefore they are less likely to fall outside the unit circle for moderate SNR values. Fig. 1(i) shows the SVD method but with the data used in the forward direction [unlike (2)]. This should be compared with Fig. 1(c) to appreciate the reduction in perturbations achieved by SVD.

Experiment 2

In this experiment we compute the estimates of the pole damping factors ($\hat{\alpha}_k$'s) and pole frequencies ($\hat{f}_k$'s) for 500 independent trials at different SNR values using $N = 25$ data samples in each case. Both cases of one ($M = 1$) and two ($M = 2$) exponentially damped signals are used. The SNR is defined the same way as in experiment 1. We emphasize that it is the SNR at the peak value of one signal component. The sample variances of the estimates were calculated and the results are plotted in Fig. 2(a)-(d). For each trial, the polynomial $B(z)$ was found using the SVD method, (4), and the M zeros outside the unit circle were used to find the estimate of the parameters by reflecting them inside the circle. If more than M zeros occurred outside the circle, which happens usually at low SNR, it generally results in large estimation error, and a threshold is said to occur. Fig. 2(a)-(d) also shows the appropriate CR bounds which are calculated using formulas derived in the Appendix. The sample variance of the estimates in all cases had a broad minimum some where in between $L = 10$ and 20. The bias in the frequency parameters was always small in comparison to the standard deviation above the threshold SNR. But the bias in the damping factors were significant at low SNR values (<20 dB) especially when the damping factor is large ($\alpha_1 = 0.2$). A relatively large value of L equal to 18 was chosen to keep this bias small. Fig. 3(a) shows a case where the bias in the damping factor estimates is conspicuous. This type of radial bias is common in LP methods [17] and is due to the noise variance contributing significantly to the singular value σ_1 (or the

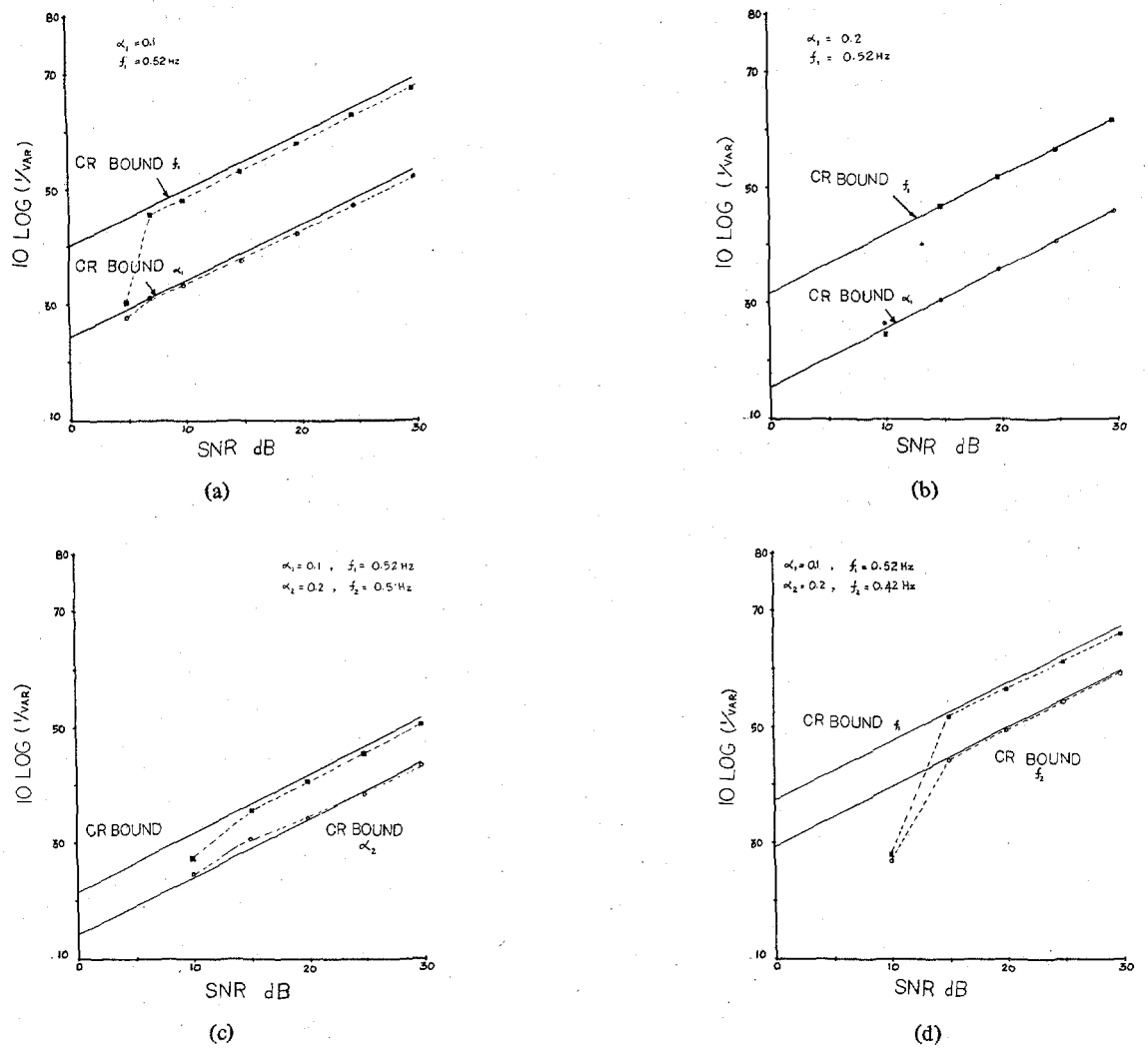

Fig. 2. (a) The sample variance of estimates of α_1 and f_1 are compared to the appropriate Cramer–Rao bounds. The data were generated using the formula $y(n) = e^{s_1 n} + w(n)$, $n = 0, 1, \cdots, 24$. $s_1 = -0.1 + j2\pi(0.52)$. 500 trials were performed. The threshold occurred at about 8 dB. L was chosen 18. (b) The details are same as in (a) except $s_1 = -0.2 + j2\pi(0.52)$. The threshold now occurred at about 15 dB. (c) The sample variance of $\hat{\alpha}_1, \hat{\alpha}_2$ compared to the appropriate CR bounds. The data (same as in Fig. 1) were generated using the formula $y(n) = e^{s_1 n} + e^{s_1 n} + w(n)$, $n = 0, 1, \cdots, 24$. $s_1 = -0.1 + j2\pi(0.52)$ and $s_2 = -0.2 + j2\pi(0.42)$. 500 trials were performed. $L = 18, M = 2$. The threshold occurred at about 11 dB. The bias was significant in estimates of below 20 dB SNR. (d) The sample variance of the estimates of f_1 and f_2 are compared to the CR bounds. The details are the same as in (c).

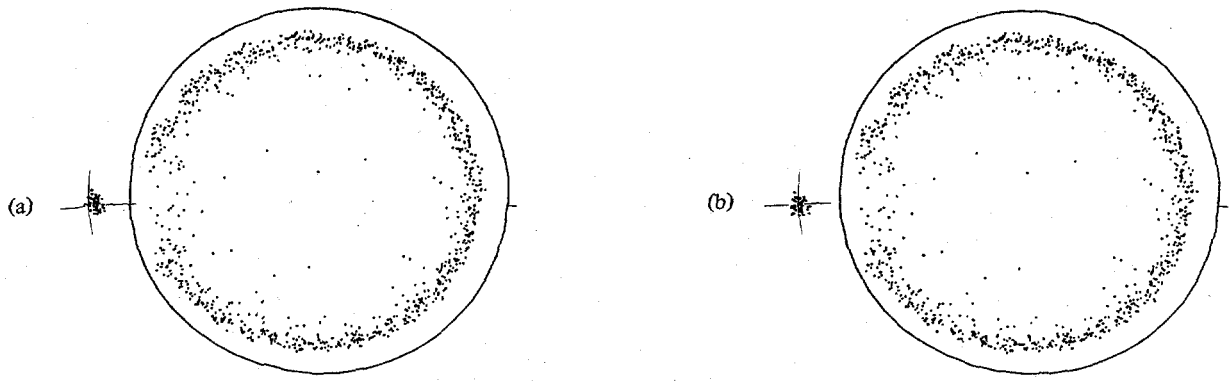

Fig. 3. (a) Examples of radial bias in the signal zero cluster. Data: $y(n) = e^{s_1 n} + w(n)$, $s_1 = -0.2 + j2\pi(0.52)$, $N = 25$, $L = 18$, SNR = 15 dB. The zeros of the prediction-error filter $B(z)$ for 40 independent data blocks are superimposed. (b) The reduction in bias is achieved by using a noise compensation procedure. From the largest singular value σ_1, the number $(\sigma_2 + \sigma_3 + \sigma_4 + \sigma_5 + \sigma_6 + \sigma_7)/6$ is subtracted before computing b in (4). The same data sets as in (a) are used.

largest eigenvalue of $A^{\dagger}A$.) This can be compensated for by subtracting from the M (equal to one in this case) largest singular values of A, a positive number equal to the average of the rest of the singular values $\sigma_{M+1}, \sigma_{M+2}, \cdots, \sigma_{N-L}$. Fig. 3(b) shows the situation when the bias compensation is applied. In Table I the bias in the estimates of α_1 before and after this modification is computed and tabulated for different SNR values. Similar results are true for the two signal case as well.

It is seen from Fig. 2(a)-(d), that the sample variance of the parameter estimates is quite close to the appropriate CR bounds. Interestingly, the larger the damping factor the closer is the sample variance to the CR bound. However, the threshold SNR is larger for the signal with larger damping factor as can be expected. The threshold occurs, generally, at a SNR below which a sudden change in the size of the singular values σ_M and σ_{M+1} is not obvious. For quantitative interpretation of the results in this experiment, it is important to recall that the SNR is the peak SNR of one of the signal components, and not an average SNR over the data interval.

Experiment 3

In this experiment the noise corrupted samples of an impulse response of a filter having poles and zeros are used as the data. This example is drawn from [8].

The pole-zero filter has a transfer function $H(z)$ whose magnitude is shown in Fig. 4.

$$H(z) = \frac{N(z)}{D(z)}$$

$$N(z) = 1 + \sum_{k=1}^{2} n(k)\, z^{-k}$$

$$D(z) = 1 + \sum_{k=1}^{10} d(k)\, z^{-k} \tag{6}$$

where $n(1) = -1.414$, $n(2) = 1.0$, and $d(1)$ to $d(10)$ are equal to -3.301, 7.226, -11.637, 14.728, -15.636, 13.588, -9.938, 5.774, -2.477, 0.735, respectively. The filter has two zeros on the unit circle. First, forty real valued samples of the impulse response in additive real valued white Gaussian noise are observed. The SNR is defined as

$$10 \log \left(\sum_{n=0}^{39} |h(n)|^2 / N\sigma \right)^2$$

where σ^2 is the variance of a noise sample and $h(n)$ is the noiseless impulse response sequence. The procedure outlined in Section II is applied to these 40 ($=N$) samples. L is chosen 20 and M is, of course 10. The twenty zeros of $B(z)$ [whose coefficients are found as in (4)] for 20 independent trials are plotted in Fig. 5. The estimates of the pole locations of $H(z)$, $e^{\hat{s}_k}$, $k = 1, 2, \cdots, M$ are obtained from the ten zeros of $B(z)$ that fall outside the circle. The estimate of the denominator polynomial is

$$\hat{D}(z) = \prod_{k=1}^{10} (1 - z^{-1} e^{\hat{s}_k}).$$

TABLE I
THE BIAS IN ESTIMATES OF α_1, BEFORE AND AFTER THE NOISE COMPENSATION SHOWN IN FIG. 3(b) IS CALCULATED AND TABULATED FOR DIFFERENT SNR VALUES

SNR dB	Bias $(\alpha_1 - \hat{\alpha}_1)$ Unmodified	Bias $(\alpha_1 - \hat{\alpha}_1)$ Modified
30	0.3574×10^{-3}	0.3538×10^{-3}
25	0.3059×10^{-2}	0.6849×10^{-3}
20	0.9733×10^{-2}	0.1279×10^{-2}
15	0.3124×10^{-1}	0.1132×10^{-2}

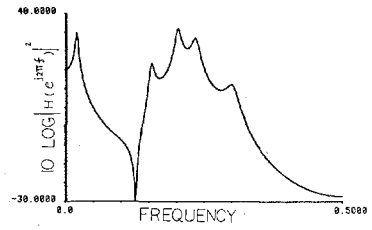

Fig. 4. Magnitude of the transfer function of the filter $H(z)$ defined in (6).

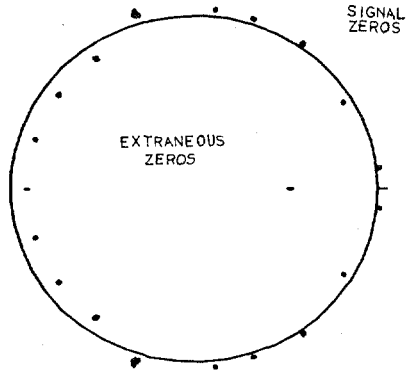

Fig. 5. The zeros of $B(z)$ obtained in twenty independent trials are superimposed. $L = 20$, $M = 10$. The first forty samples of the noise corrupted impulse response of $H(z)$ are used as the data. SNR = 25 dB.

Subsequently, the numerator polynomial estimate $\hat{N}(z)$ is obtained by using Shanks' method [13], [18]. Shanks' method is as follows. A sequence $f(n)$ is generated, the Z transform of which is $F(z)$

$$F(z) = \frac{1}{\hat{D}(z)}. \tag{7}$$

Then the numerator polynomial estimate $\hat{N}(z)$ is found by minimizing the error

$$E = \sum_{m=0}^{39} \left| y(m) - \sum_{k=0}^{2} \hat{n}(k)\, f(m-k) \right|^2 \tag{8}$$

by finding $\hat{n}(k)$, $k = 0, 1, 2$. This is a linear least square problem. Fig. 6 shows the magnitude of the transfer function of the reconstructed filter $\hat{H}(z) = \hat{N}(z)/\hat{D}(z)$ for two different SNR values. It is interesting to note that in Fig. 6(b), the 5th (highest frequency) formant exhibits sharper peaks compared to Fig. 4. This formant corresponds to the largest pole damping factor. The sharper peaks are caused by the radial bias in the corresponding zeros of $B(z)$ similar to that in Fig. 3(a).

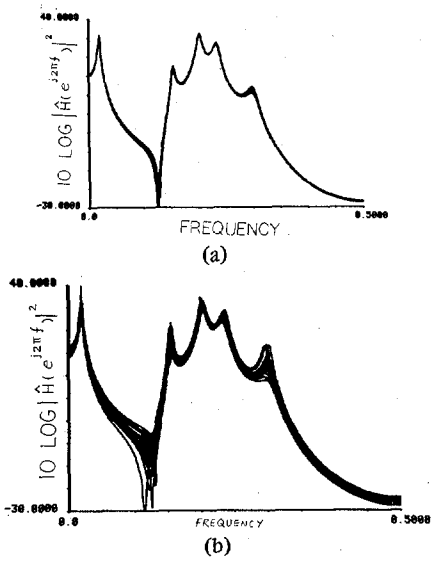

Fig. 6. The magnitude of the transfer function of the reconstructed filter $\hat{H}(z)$. The zeros of the filter were determined using Shank's method. (a) SNR = 40 dB. (b) SNR = 25 dB.

This contrasts with the traditional LP methods where additive noise always produces flatter spectral peaks [17].

IV. Conclusions

A method of estimating the parameters of exponentially damped signals in noise based on singular value decomposition and backward linear prediction is described. The method is shown to give more accurate estimates of the parameters than traditional methods. Further comparisons are given in [21].

Appendix

In this appendix we shall describe an outline of the calculation of the CR bounds for the parameters of one ($M = 1$) and two ($M = 2$) exponentially damped signals in white Gaussian noise. At high SNR values our parameter estimates are essentially unbiased. The CR bounds, in that case, provide a lower bound on the variance of the parameter estimates. The data samples $y(n)$ are given by the formula

$$y(n) = \sum_{k=1}^{m} c_k e^{j\theta_k} e^{-\alpha_k n + j2\pi f_k n} + w(n), \quad n = 0, 1, \cdots, N-1 \tag{A1}$$

where $w(n)$ is a white Gaussian sequence. Sampling interval is assumed one second. Each real and imaginary part of $w(n)$ has a variance σ^2. Comparing (1) with the above formula, we see that $a_k = c_k e^{j\theta_k}$ and $s_k = -\alpha_k + j2\pi f_k$, $k = 1, 2, \cdots, M$. For convenience we shall relabel the $4M$ parameters as follows:

$$\begin{aligned} \beta_{4(k-1)} &= f_k \\ \beta_{4(k-1)+1} &= c_k \qquad k = 1, 2, \cdots, M \\ \beta_{4(k-1)+2} &= \theta_k \\ \beta_{4(k-1)+3} &= \alpha_k \end{aligned} \tag{A2}$$

$$y = (y(0), y(1), \cdots, y(N-1))^T \tag{A3}$$

$$\beta = (\beta_0, \beta_1, \beta_2, \cdots, \beta_{4(m-1)+3})^T. \tag{A4}$$

The probability density function for the data vector conditioned on the unknown parameter vector can be written as

$$P(y|\beta) = (2\pi\sigma^2)^{-N} \exp\left[\frac{-1}{2\sigma^2} \sum_{n=0}^{N-1} \cdot \left| y(n) - \sum_{k=1}^{M} c_k e^{j\theta_k} e^{(-\alpha_k + j2\pi f_k)n} \right|^2\right] \tag{A5}$$

The CR bound states [19] that for any unbiased estimate of $\hat{\beta}_i$ of β_i.

$$\text{Var}(\hat{\beta}_i) \geq [J^{-1}(\beta)]_{ii} \tag{A6}$$

where $J(\beta)$ is the $(4M \times 4M)$ Fisher information matrix with elements [19]

$$[J(\beta)]_{ij} = -E\left[\frac{\partial^2}{\partial\beta_i \partial\beta_j} \ln P(y|\beta)\Big|_{\beta}\right]. \tag{A7}$$

We shall consider the cases $M = 1$ and 2. The derivation follows that of Rife and Boorstyn [20] who derived similar expressions for the case of sinusoidal signals.

Case 1: $M = 1$

Using formula (A7) above we can write down the Fisher matrix for this case as follows:

$$J = J_{ii} = \begin{bmatrix} c_i^2 p_{i2} & 0 & c_i^2 p_{i1} & 0 \\ & p_{i0} & 0 & -c_i p_{i1} \\ & & c_i^2 p_{i0} & 0 \\ \text{symmetric} & & & c_i^2 p_{i2} \end{bmatrix} \quad i = 1 \tag{A8}$$

where

$$P_{ij} = \sum_{n=0}^{N-1} n^j e^{-2\alpha_i n}, \quad j = 0, 1, 2. \tag{A9}$$

Inverting J analytically, we find the diagonal terms of J which provide the bounds on the variance for the parameters f_1 and α_1 as follows:

$$\text{Var}(\hat{f}_1) \geq \frac{\sigma^2}{4\pi^2 c_1^2} \frac{p_{10}}{p_{10}p_{12} - p_{11}^2} \tag{A10}$$

$$\text{Var}(\alpha_1) \geq \frac{\sigma^2}{c_1^2} \frac{p_{10}}{p_{10}p_{12} - p_{11}^2}. \tag{A11}$$

Case 2: $M = 2$.

In this case J is an (8×8) matrix. We partition J as follows:

$$J = \begin{bmatrix} J_{11} & J_{12} \\ J_{12}^T & J_{22} \end{bmatrix}$$

where the matrices J_{11} and J_{22} are given by (A8) and (A9) with $i = 1$ and $i = 2$, respectively. J_{12} is a matrix exhibiting the interaction between the parameters of the two signals.

$$J_{12} = \begin{bmatrix} c_1c_2\gamma_2 & c_1q_1 & c_1c_2\gamma_1 & -c_1c_2q_2 \\ -c_2q_1 & \gamma_0 & -c_2q_0 & -c_2\gamma_1 \\ c_1c_2\gamma_1 & c_1q_0 & c_1c_2\gamma_0 & -c_1c_2q_1 \\ c_1c_2q_2 & -c_1\gamma_1 & c_1c_2q_1 & c_1c_2\gamma_2 \end{bmatrix} \tag{A13}$$

where

$$\left.\begin{aligned}\gamma_j &= \sum_{n=0}^{N-1} e^{-(\alpha_1+\alpha_2)n}\, n^j \cos\Delta_n \\ q_j &= \sum_{n=0}^{N-1} e^{-(\alpha_1+\alpha_2)n}\, n^j \sin\Delta_n\end{aligned}\right\} \quad j = 0, 1, 2. \tag{A14}$$

$$\Delta_n = 2\pi(f_2 - f_1)\, n + \theta_2 - \theta_1$$

The matrix J was inverted using a machine to compute the CR bounds of the parameters f_1, f_2, α_1, and α_2 which are used in Fig. 2(c) and (d). Interestingly, due to the special structure of J_{12}, it turns out that the diagonal elements of J^{-1} are independent of $\theta_2 - \theta_1$.

References

[1] D. W. Tufts and R. Kumaresan, "Improved spectral resolution II," in *Proc. ICASSP 80*, Apr. 1980, pp. 592-597.

[2] R. Kumaresan and D. W. Tufts, "Data adaptive principal component signal processing," in *Proc. 19th IEEE Int. Conf. Decision and Contr.*, Feb. 10-12, 1981, Albuquerque, NM.

[3] —, "Improved spectral resolution III: Efficient realization," *Proc. IEEE*, vol. 68, Oct. 1980.

[4] D. W. Tufts, R. Kumaresan, and I. Kirsteins, "Data adaptive signal estimation," accepted for publication in *Proc. IEEE*.

[5] D. W. Tufts and R. Kumaresan, "Frequency estimation of multiple sinusoids: Making linear prediction perform like maximum likelihood," accepted for publication in *IEEE Trans. Acoust., Speech, Signal Processing.*

[6] —, "Singular value decomposition and frequency estimation by linear prediction," accepted for publication in *IEEE Trans. Acoust., Speech, Signal Processing*.

[7] R. Kumaresan and D. W. Tufts, "Singular value decomposition and spectral analysis," in *Proc. 1st Acoust., Speech, Signal Processing workshop on spectral estimation*, Aug. 17 and 18, 1981, Hamilton, Ont., Canada, pp. 6.4.1-6.4.12.

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

[9] T. J. Ulrych and R. W. Clayton, "Time series modelling and maximum entropy," *Phys. Earth and Planetary Interiors*, vol. 12, pp. 188-200, Aug. 1976.

[10] A. H. Nuttall, "Spectral analysis of a univariate process with bad data points via maximum entropy and linear predictive techniques," in *NUSC Scientific and Engineering Studies, Spectral Estimation, NUSC*, New London, CT, Mar. 1976.

[11] F. B. Hildebrand, *Introduction to Numerical Analysis*. New York: McGraw-Hill, 1956.

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

[13] J. L. Shanks, "Recursion filters for digital processing," *Geophysics*, vol. 32, pp. 33-51, 1967.

[14] R. Kumaresan, "On the zeros of the linear prediction error filter for deterministic signals," submitted for publication to *IEEE Trans. Acoust., Speech, Signal Processing.*

[15] J. D. Markel and A. H. Gray, *Linear Prediction of Speech*. New York, Springer-Verlag, 1976.

[16] C. L. Lawson and R. J. Hanson, *Solving Least Squares Problems*. Englewood Cliffs, NJ: Prentice-Hall, 1974.

[17] S. M. Kay, "The effects of noise on the autoregressive spectral estimator," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-27, Oct. 1979.

[18] G. E. Kopec, A. V. Oppenheim, and J. M. Tribolet, "Speech analysis by homomorphic prediction," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-27, pp. 40-49, Feb. 1977.

[19] H. J. Larson, *Introduction To The Theory of Statistics*. New York: Wiley, 1973.

[20] D. C. Rife and R. R. Boorstyn, "Multiple tone parameter estimation from discrete time observations," *Bell Syst. Tech. J.*, pp. 1389-1410, Nov. 1976.

[21] R. Kumaresan and D. W. Tufts, "Accurate parameter estimation of noisy speech-like signals," presented at the IEEE Int. Conf. on Acoust., Speech, Signal Processing, May 3-5, 1982, Paris, France.

Ramdas Kumaresan received the Bachelor of Engineering degree with honors from the University of Madras, India in 1971, and the M.S. and Ph.D. degrees from the University of Rhode Island, Kingston, in 1979 and 1982, respectively.

From 1972 to 1977 he was a member of the technical staff with Bharat Electronics Limited, India where he was involved in designing radar equipment. From 1977 to 1979 he was a Research Assistant at the University of Rhode Island. From 1980 to 1982 he was a Research Associate. He is presently a member of the Technical Staff of GTE Business Communication Systems, McLean, VA. His research interests include signal processing, spectral analysis, and applied mathematics.

Donald W. Tufts (S'58-M'61-SM'78-F'82) was born on March 5, 1933 in Yonkers, N.Y. He received the B.A. degree in mathematics from Williams College, Williamstown, MA, in 1955. In 1957, 1958, and 1960 he received the S.B., S.M., and Sc.D. degrees from the Massachusetts Institute of Technology, Cambridge, where he was a Bell Telephone Laboratories Fellow.

From 1960 to 1967 he was at Harvard University, first as Research Fellow and Lecturer and later as Assistant Professor of Applied Mathematics. From 1967 to the present he has been a Professor of Electrical Engineering at the University of Rhode Island, Kingston. He has been a consultant to Bell Telephone Laboratories, Sanders Associates Inc., and to other companies.