FORMANT ESTIMATION USING SINGULAR VALUE DECOMPOSITION

Invited Paper

Raymond E. Slyh and Randolph L. Moses
Department of Electrical Engineering
The Ohio State University
2015 Neil Avenue

Abstract

This paper considers the estimation of a small number of damped complex exponentials from data which contain a larger set of damped exponentials and noise. The use of total least squares methods results in biased estimates of the high energy damped exponentials due to the presence of the extra modes. We discuss the reasons for this bias and ways to reduce it.

I. Introduction

There are many applications in speech processing where one wishes to estimate the formant frequencies, as well as the associated bandwidths and modal energies, from a short speech record. Of the many methods available for this estimation problem, one of the most successful is the total least squares (TLS) method, which consists of backward prediction coupled with the singular value decomposition (SVD) [1,2,3,4]. Typical derivations of formant estimation algorithms assume that the speech signal can be modeled as a sum of damped exponentials in additive white noise [1]. It is of interest to estimate the strongest three or four formants. On the other hand, the speech signal may contain five or more formants; these additional modes are not well modeled as white noise. The question arises as to what effect these additional formants have on the estimation of the dominant formants.

In this paper, we consider the problem of estimating the highest energy modes from noisy data using total least squares techniques [2]. The problem to be considered can be formulated as follows. Suppose that we have N data points of a noisy complex signal, $\{y_k\}_{k=1}^{N}$, which can be modeled as

$$y_k = \sum_{i=1}^{M} A_i \rho_i^k + n_k, \tag{1}$$

where $\{n_k\}_{k=1}^{N}$ is a white noise sequence with variance ν^2, and A_i is the complex amplitude associated with the complex pole ρ_i. Here, M is the number of modes in the model. Of the M modes, we wish to accurately estimate the parameters for only the H highest-energy modes, where $H \leq M$. We refer to this as the reduced-order modeling problem.

In order to better understand the SVD used for reduced-order modeling purposes, this paper examines the relationship between the SVD and modal decompositions. The next section discusses the equivalence of the SVD and modal decompositions, while the third section shows the results of some simulations regarding the choosing of the number of singular values to keep in the TLS algorithm for reduced-order modeling purposes.

This work was supported in part by the Office of Naval Research under contract number N00014-86-K-0202.

II. Modal Decomposition and SVD Equivalence

In order to characterize the behavior of the TLS algorithm used for reduced-order modeling purposes, we shall first look at the SVD. Before performing the linear prediction in the TLS algorithm, the $(N-L) \times (L+1)$ Hankel matrix $\mathbf{Y}$ is formed as

$$\mathbf{Y} = \begin{bmatrix} y_1 & y_2 & \cdots & y_{L+1} \\ y_2 & y_3 & \cdots & y_{L+2} \\ \vdots & \vdots & & \vdots \\ y_{N-L} & y_{N-L+1} & \cdots & y_N \end{bmatrix}, \tag{2}$$

where L is the linear prediction model order chosen to be greater than M. The SVD of $\mathbf{Y}$ is then formed as

$$\mathbf{Y} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^H, \tag{3}$$

where $\mathbf{U}$ and $\mathbf{V}$ are square unitary matrices, ^{H} is the complex conjugate transpose operator, and $\mathbf{\Sigma}$ is an $(N-L) \times (L+1)$ matrix such that $\mathbf{\Sigma} = \mathrm{diag}(\sigma_1, \sigma_2, \ldots, \sigma_{\min\{(N-L),(L+1)\}})$. The σ_j are the singular values of $\mathbf{Y}$; the singular values are real, nonnegative, and ordered in magnitude such that $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_{\min\{(N-L),(L+1)\}}$. To provide noise cleaning of the data, the singular values are truncated so that

$$\hat{\mathbf{Y}} = \mathbf{U}\hat{\mathbf{\Sigma}}\mathbf{V}^H, \tag{4}$$

where $\hat{\mathbf{Y}}$ is the "noise-cleaned" data matrix. The truncated matrix containing the singular values, $\hat{\mathbf{\Sigma}}$, is given by

$$\hat{\mathbf{\Sigma}} = \mathrm{diag}(\sigma_1, \sigma_2, \ldots, \sigma_R, 0, \ldots, 0), \tag{5}$$

where $R < \min\{(N-L), (L+1)\}$.

This truncation of singular values is intended to reduce the rank of $\mathbf{Y}$; however, it is desirable to know what effect this truncation has on the underlying modes. Specifically, if one truncates the singular values so that R is less than M (see equation (1)), will the "weak" (low energy) modes of the data be completely purged as well as the noise? In the context of reduced-order modeling, R is often chosen to be H, the number of desired (high energy) modes. However, it is not true in general that, by choosing $R = H$, the H highest energy modes in (1) will be estimated, as we show below.

In order to relate the SVD truncation operation with model reduction, it is useful to consider the modal decomposition of $\mathbf{Y}$.

From equations (1) and (2), the noiseless data matrix $\mathbf{Y}$ can be written in a modal decomposition form [5]

$$\mathbf{Y} = \begin{bmatrix} \rho_1 & \rho_2 & \cdots & \rho_M \\ \rho_1^2 & \rho_2^2 & \cdots & \rho_M^2 \\ \vdots & \vdots & & \vdots \\ \rho_1^{N-L} & \rho_2^{N-L} & \cdots & \rho_M^{N-L} \end{bmatrix} \times \cdots$$
$$\begin{bmatrix} A_1 & 0 & \cdots & 0 \\ 0 & A_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & A_M \end{bmatrix} \begin{bmatrix} 1 & \rho_1 & \cdots & \rho_1^L \\ 1 & \rho_2 & \cdots & \rho_2^L \\ \vdots & \vdots & & \vdots \\ 1 & \rho_M & \cdots & \rho_M^L \end{bmatrix}. \tag{6}$$

A more general version of the modal decomposition follows directly from (6):

$$\mathbf{Y} = \begin{bmatrix} \frac{\rho_1}{F_1} & \frac{\rho_2}{F_2} & \cdots & \frac{\rho_M}{F_M} \\ \frac{\rho_1^2}{F_1} & \frac{\rho_2^2}{F_2} & \cdots & \frac{\rho_M^2}{F_M} \\ \vdots & \vdots & & \vdots \\ \frac{\rho_1^{N-L}}{F_1} & \frac{\rho_2^{N-L}}{F_2} & \cdots & \frac{\rho_M^{N-L}}{F_M} \end{bmatrix} \times \cdots$$
$$\begin{bmatrix} F_1 & 0 & \cdots & 0 \\ 0 & F_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & F_M \end{bmatrix} \begin{bmatrix} A_1 & A_1\rho_1 & \cdots & A_1\rho_1^L \\ A_2 & A_2\rho_2 & \cdots & A_2\rho_2^L \\ \vdots & \vdots & & \vdots \\ A_M & A_M\rho_M & \cdots & A_M\rho_M^L \end{bmatrix} \tag{7}$$

or

$$\mathbf{Y} = \mathbf{P}\mathbf{D}\mathbf{Q}^H, \tag{8}$$

where the F_i coefficients are positive real numbers, ordered so that $F_1 \geq F_2 \geq \cdots \geq F_M \geq 0$. For example, F_i could be defined as the energy of a mode:

$$F_i = |A_i|^2 \sum_{k=1}^{N} |\rho_i|^{2k}, \text{ for } i = 1, 2, \ldots, M. \tag{9}$$

The generalized modal decomposition is useful for the following reason. Suppose that one is interested in a model for $\{y_k\}$ which contains only the R highest energy modes. This model is found directly from equations (7) and (8) by truncating the $\mathbf{D}$ matrix, i.e. by replacing $\mathbf{D}$ with

$$\hat{\mathbf{D}} = \begin{bmatrix} \mathbf{D}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}, \tag{10}$$

where $\mathbf{D}_1 = \operatorname{diag}(F_1, \ldots, F_R)$. Moreover, F_i can be defined as some other function (not necessarily energy), and other rules for model reduction could be used (for example, the R highest amplitude modes could be kept by choosing $F_i = |A_i|$ and truncating as before). Thus, the generalized modal decomposition gives a direct solution to the generalized model reduction problem.

The SVD of $\mathbf{Y}$ in (3) is, in general, not the modal decomposition in (7). For the SVD to be equivalent to a modal decomposition, it is necessary that the columns of $\mathbf{P}$ and $\mathbf{Q}$ in the modal decomposition be orthogonal; that is

$$\mathbf{P}^H\mathbf{P} = \mathbf{\Lambda}_P, \text{ and } \mathbf{Q}^H\mathbf{Q} = \mathbf{\Lambda}_Q, \tag{11}$$

where $\mathbf{\Lambda}_P$ and $\mathbf{\Lambda}_Q$ are $M \times M$ diagonal matrices with positive diagonal elements λ_{Pi} and λ_{Qi} respectively. The conditions of (11) combine to yield [6]

$$\frac{1}{F_iF_j} \sum_{k=1}^{N-L} (\rho_i^*\rho_j)^k = 0 \text{ for all } i \neq j, \tag{12}$$

and

$$\frac{A_iA_j^*}{\rho_i\rho_j^*} \sum_{k=1}^{L+1} (\rho_i\rho_j^*)^k = 0 \text{ for all } i \neq j. \tag{13}$$

Conditions (12) and (13) are satisfied only if, for all $i \neq j$,

$$\left\{ \begin{array}{l} |\rho_i|\,|\rho_j| = 1 \\ (\theta_j - \theta_i) = \dfrac{2\pi k}{N-L} \text{ for some integer } k \neq 0 \end{array} \right\}, \tag{14}$$

and

$$\left\{ \begin{array}{l} |\rho_i|\,|\rho_j| = 1 \\ (\theta_i - \theta_j) = \dfrac{2\pi n}{L+1} \text{ for some integer } n \neq 0 \end{array} \right\}, \tag{15}$$

where θ_i denotes the phase angle of the pole ρ_i. Note that these conditions are very restrictive. For example, if $M \geq 3$, these conditions are satisfied only if $|\rho_i| = 1$ for all i and the angle restrictions of (14) and (15) are met. These constraints will almost never be satisfied in practice. Thus, the SVD will almost never equal a modal decomposition in practice, so truncation of singular values will not completely eliminate weaker modes. However, these conditions are useful in determining how closely a SVD truncation approximates a modal truncation. For example, if the modes of the data are nearly orthogonal in the sense that the left hand sides of (12) and (13) are nearly zero, then a SVD will be close to a modal decomposition.

Assume that the conditions of (11) are met, and let

$$\mathbf{U}_1 = \mathbf{P}\mathbf{\Lambda}_P^{-1/2}, \text{ and } \mathbf{V}_1 = \mathbf{Q}\mathbf{\Lambda}_Q^{-1/2}. \tag{16}$$

Define $\mathbf{U}_2$ and $\mathbf{V}_2$ so that $[\mathbf{U}_1|\mathbf{U}_2]$ and $[\mathbf{V}_1|\mathbf{V}_2]$ are unitary matrices. Then, from (3), (7), and (16), it follows that

$$\mathbf{Y} = \mathbf{U}_1\mathbf{\Sigma}_1\mathbf{V}_1^H = \begin{bmatrix} \mathbf{U}_1 & \mathbf{U}_2 \end{bmatrix} \begin{bmatrix} \mathbf{\Sigma}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{V}_1^H \\ \mathbf{V}_2^H \end{bmatrix} \tag{17}$$

is a singular value decomposition of $\mathbf{Y}$, where $\mathbf{\Sigma}_1 \triangleq \mathbf{\Lambda}_P^{1/2}\mathbf{D}\mathbf{\Lambda}_Q^{1/2}$. Thus, if the conditions of (11) are met, then the modal decomposition and SVD are equivalent to within a scaling factor. In this case, the M diagonal elements of $\mathbf{\Sigma}_1$ are given by

$$\begin{aligned} \sigma_i &= \frac{|A_i|}{|\rho_i|} \left(\sum_{k=1}^{N-L} |\rho_i|^{2k} \right)^{1/2} \left(\sum_{k=1}^{L+1} |\rho_i|^{2k} \right)^{1/2} \\ &= \frac{1}{|A_i\rho_i|} \left(\sum_{k=1}^{N-L} |A_i|^2 |\rho_i|^{2k} \right)^{1/2} \left(\sum_{k=1}^{L+1} |A_i|^2 |\rho_i|^{2k} \right)^{1/2} \end{aligned} \tag{18}$$

It can be seen that σ_i is a weighted geometric mean of the energy of the i^{th} mode for two different observation intervals. Thus, truncating singular values removes the lowest energy modes in (1). Of course, this result holds only when the conditions (12) and (13) are satisfied exactly. When (12) and (13) are nearly satisfied, singular-value truncation removes the low energy modes only approximately; see [6] for more details.

As an illustration of conditions (14) and (15), consider the following case. Suppose that

$$y_k = \cos(\frac{\pi k}{11}) + .9\cos(\frac{3\pi k}{11}), \tag{19}$$

and that we use a prediction order, L, of 10. If $N = 26$, the four poles do not satisfy the conditions for the SVD to be equivalent to a modal decomposition; however, if $N = 21$, the conditions are satisfied. Thus, if we keep only two singular values, we should get biased results using 26 data points; whereas, we should get unbiased results using only 21 data points. This is verified by Figures 1 and 2, which show the results of keeping two singular values in the TLS algorithm using 21 and 26 data points, respectively. These two plots assume a sampling rate of 10 KHz, so that the two true spectral peaks (denoted by the dashed lines) lie at 454.5 Hz and 1363.6 Hz. The solid lines indicate the estimated spectra obtained by keeping only two singular values in the TLS algorithm.

III. The TLS Algorithm and Reduced-Order Modeling

In the previous section, we have seen that the SVD is rarely a modal decomposition. This leads to biased estimates when one keeps fewer singular values than the number of modes in the data. This would appear to indicate that, to reduce the estimate bias, one should keep as many singular values as possible; however, this must be balanced against the corresponding increase in estimate variance which will accompany such a scheme. The result is a bias/variance tradeoff [6]. We will investigate this tradeoff in this section by the means of simulation results.

For the simulations of this section, a data set has been created to have the general characteristics of speech data. This data set is a sum of eight damped complex exponentials (equivalently, four damped sinusoids). The eight complex poles and their corresponding amplitudes are as shown in Table 1; the corresponding frequencies are also listed. Note that the poles are listed in order of decreasing energy.

Suppose that we wish to estimate the two highest energy sinusoids. To do this [6], we keep some number of singular values

Table 1: Data Set Parameter Values

A_i	$\mid \rho_i \mid$	$\angle \rho_i$ in rad/π	Frequency (Hz)
1.0000	0.9709	± 0.1300	650.3
0.8000	0.9717	± 0.2152	1075.7
0.6000	0.9669	± 0.4926	2463.1
0.4000	0.9395	± 0.7117	3558.3

in the TLS algorithm and estimate the poles for the data. Using these poles, we estimate the corresponding amplitudes of the poles. Finally, we calculate the energies of all the modes, keep the higher energy modes (keep four pole/amplitude pairs), and calculate the spectrum using the chosen modes. We wish to examine the results of keeping various numbers of singular values in the algorithm. For all of the simulation results of this section, $N = 20$ and $L = 10$.

Figures 3 and 4 show the results of keeping four and eight singular values, respectively, in the TLS algorithm and then choosing the two highest energy damped sinusoids. For both figures, the signal to noise ratio (SNR) is 26.2 db. Also, for both figures, there are ten overlayed simulations (denoted by the solid lines); the dashed lines indicate the exact spectra using all four damped sinusoids. We see that keeping four singular values in this case results in choosing the first and third highest energy modes, while keeping eight singular values results in choosing the two highest energy modes. Thus, for low noise (relative to the powers of the lower energy modes), one should keep a larger number of singular values than the number of poles that we wish to examine; this decreases the bias resulting from the fact that the SVD is not a modal decomposition [6].

Figures 5 and 6 show the results of keeping four and eight singular values, respectively, in the TLS algorithm when the SNR is 16.2 db. The solid and dashed lines have the same meaning as before. Again, keeping four singular values in this case results in choosing the first and third highest energy modes; however, keeping eight singular values sometimes results in choosing the two highest energy modes and sometimes results in choosing the first and third modes. The best choice for the number of singular values to keep in this case depends upon the application. In formant tracking, for example, one does not want the estimated formants to "jump around" from the second to the third and back again; however, one generally wants to examine the highest energy modes. In this case, one must compromise. Keeping a small number of singular values results in more bias but less variance, while the opposite is true for keeping a larger number of singular values [6].

The simulation results of this section indicate that if one wishes to examine only the highest energy modes of some data set, one can use the TLS algorithm keeping the proper number of singular values for the noise scenario; after calculating the associated amplitudes using a least squares technique, one need only calculate the modal energies and keep only those modes with the highest energy. This assumes that the noise power is relatively low; if it is not, then the resulting estimates may not be of the highest energy modes.

IV. Conclusions

We have seen that the singular value decomposition is rarely a modal decomposition when the data is modeled as in equation (1). The elimination of singular values, therefore, is not equivalent to the elimination of modes from the data. The result is that it is not always possible, by keeping H singular values, to accurately estimate the parameters for H modes when there are M modes in the data and $H \leq M$.

We have examined the bias and variance performance of the total least squares (TLS) algorithm versus the number of singular values, R, kept in the algorithm in the context of reduced-order modeling. The simulations which have been considered indicate that there is a bias/variance tradeoff in the TLS algorithm with respect to the value of R. Choosing R to be small results in estimates with more bias, but with less variance, than estimates generated by the TLS algorithm using a larger value of R. Thus, the proper number of singular values to keep in the TLS algorithm for reduced-order modeling purposes is a function of the noise power relative to the powers of the unmodeled (undesired or low energy) modes. When the noise power is low compared to the powers of the undesired modes, one should generally choose R to be large. When the relative noise power is high, one should choose R depending upon the particular application; for many applications, R will be chosen to be small. After estimating the poles and their associated amplitudes, one should then determine the modal energies and keep only those modes with a high enough energy.

V. References

[1] S. Parthasarathy and D. W. Tufts, "Excitation-synchronous

modeling of voiced speech," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. ASSP–35, no. 9, pp. 1241–1249, September 1987.

[2] M. A. Rahman and K. 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, October 1987.

[3] V. C. Klema and A. J. Laub, "The singular value decomposition: Its computation and some applications," *IEEE Transactions on Automatic Control*, vol. AC–25, no. 2, pp. 164–176, April 1980.

[4] G. H. Golub and C. F. V. Loan, *Matrix Computations*. Baltimore: Johns Hopkins University Press, 1983.

[5] K. Glover, "An approximate realization algorithm which directly identifies the system poles," in *Proceedings of the 7th IFAC World Congress*, pp. 1789–1796, Helsinki, 1978.

[6] R. E. Slyh, "Total least squares techniques and reduced-order modeling with application to radar target identification," M.S. Thesis, The Ohio State University, August 1989.

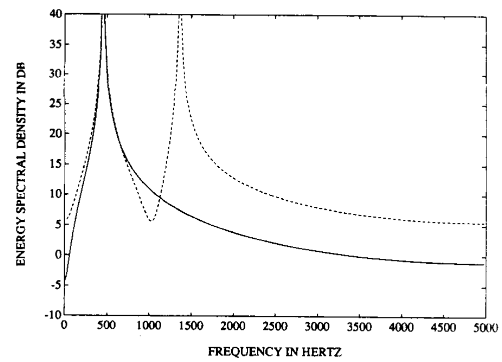

Figure 1: Energy spectrum of the two sinusoids using $R = 2$, $L = 10$, and $N = 21$.

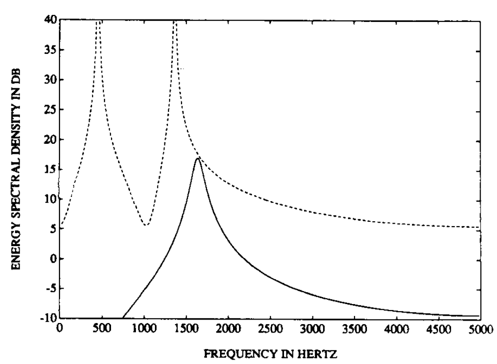

Figure 2: Energy spectrum of the two sinusoids using $R = 2$, $L = 10$, and $N = 26$.

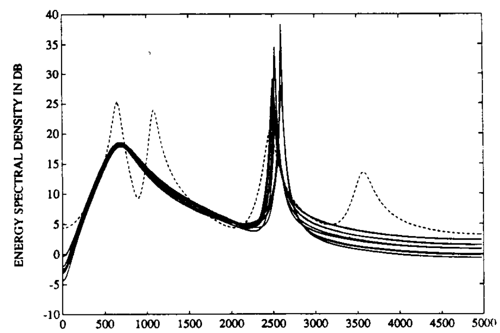

Figure 3: Energy spectra of the multisinusoid data using $R = 4$, $L = 10$, $N = 20$, and a SNR of 26.2 db.

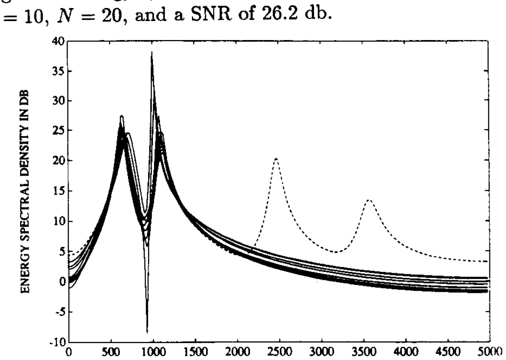

Figure 4: Energy spectra of the multisinusoid data using $R = 8$, $L = 10$, $N = 20$, and a SNR of 26.2 db.

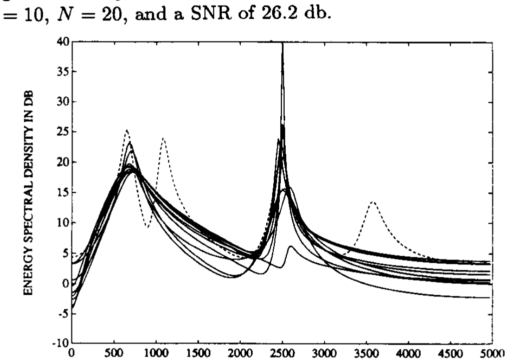

Figure 5: Energy spectra of the multisinusoid data using $R = 4$, $L = 10$, $N = 20$, and a SNR of 16.2 db.

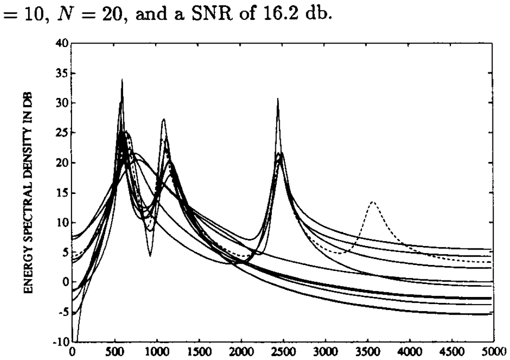

Figure 6: Energy spectra of the multisinusoid data using $R = 8$, $L = 10$, $N = 20$, and a SNR of 16.2 db.