MODE-Type Algorithm for Estimating Damped, Undamped or Explosive Modes

Mats Cedervall, Petre Stoica

Systems and Control Group
Uppsala University
Uppsala, Sweden S-75103

Randolph Moses*

Electrical Engineering
Ohio State University
Columbus, OH 43210

Abstract

We propose a new algorithm for estimating the parameters of damped, undamped or explosive sinusoidal processes. The algorithm resembles the MODE algorithm which is commonly used for direction of arrival estimation in the array signal processing field. The algorithm is asymptotically (for high SNR) optimal. Nevertheless it is computationally simple and easy to implement. Numerical examples are included to illustrate the performance of the proposed method.

1 Introduction

In this paper we consider the problem of estimating the parameters of a sum of (complex) exponential modes from noisy data. The modes may be exponentially damped, undamped, or exponentially explosive; that is, the mode "pole" locations may lie inside, on, or outside the unit circle. We focus on subspace-based estimation methods, because they provide the dual advantages of computational efficiency and good statistical properties.

A number of subspace or singular-value-based methods have been developed for estimating parameters in exponential data, including [3, 6, 7, 8]. For some of these methods, the modes are allowed to be damped or undamped, but not exponentially explosive. All of these methods exploit the low-rank property of the Hankel data matrix to obtain accurate parameter estimates in many cases. The statistical properties of many of these algorithms have been studied and compared to the Cramér-Rao bound; near optimal accuracy can be achieved by many of the algorithms (asymptotically in SNR, at least), but at the cost of using large dimensional matrices. The high matrix dimensions impose a computational burden in singular value decomposition steps.

*This work performed while on sabbatical leave at the Systems and Control Group, Uppsala University, and supported in part by a grant from the Swedish Institute.

In this paper we develop a somewhat different estimator for the exponential modes. We adopt a model used widely in sensor array processing, and apply a modification of the so-called Method Of Direction Estimation (MODE) algorithm [11]. The MODE algorithm, like those in [3, 6, 7, 8], exploits low-rank structure in higher-dimensional matrices. However, the MODE algorithm has the advantage of applying optimal weighting in the estimation problem to produce estimates which have asymptotically (as SNR $\rightarrow \infty$) minimum variance. As a result, one can use smaller matrix sizes than in the unweighted algorithms, thus reducing the computational costs related to these large matrices. The optimal weights are dependent on the parameters to be estimated; however, for high SNR, a two-step algorithm can be used to achieve (asymptotically) optimal performance at modest computational cost.

This problem has importance in a number of applications, including speech modeling [5], electrocardiogram signal modeling [4], and radar scattering analysis from stepped frequency measurements [1]. In all of these applications, both damped modes and explosive modes may arise. For example, in radar scattering, damped or explosive modes can arise because the frequency response of different scattering centers may be decreasing or increasing as a function of frequency, respectively.

2 Problem formulation

Let

$$\tilde{y}(t) = \sum_{k=1}^{n} \alpha_k \rho_k^t + \tilde{e}(t), \; t = 1, 2, \ldots N \qquad (1)$$

be the equation describing the observed signal. In (1) $\alpha_k \in \mathcal{C}$, $\tilde{e}(t)$ is a circularly symmetric Gaussian distributed noise with variance σ^2 and n is given. The number of data samples N is typically small, and the

1058-6393/96 $5.00 © 1996 IEEE
Proceedings of ASILOMAR-29

signal to noise ratio (SNR) is usually assumed to be high. The SNR of the k^{th} component in (1) is defined as follows:

$$\mathrm{SNR}_k = 10\log_{10}\left(\frac{E_k}{N\sigma^2}\right) \text{ [dB]} \tag{2}$$

where E_k is the total energy of the k^{th} mode:

$$E_k = |\alpha_k|^2 \sum_{l=0}^{N-1} |\rho_k|^{2l} = |\alpha_k|^2 \begin{cases} N \text{ if } |\rho_k| = 1 \\ \frac{1-|\rho_k|^{2N}}{1-|\rho_k|^2} \text{ otherwise} \end{cases}$$

The problem of interest is to estimate $\{\rho_k\}$ (and perhaps $\{\alpha_k\}$ as well, which is an easy task once $\{\rho_k\}$ has been obtained).

3 Solution Using MODE Estimators

In this section we present a method for estimating $\{\rho_k\}$ based on the Method Of Direction Estimation (MODE) [11]. The MODE procedure is attractive because it is computationally simple, yet yields asymptotically (in SNR) efficient parameter estimates.

To apply MODE estimation techniques to the above problem, form a vector $y(t)$ of the measured signal $\tilde{y}(t)$:

$$y(t) = \begin{bmatrix} \tilde{y}(t) \\ \vdots \\ \tilde{y}(t+m-1) \end{bmatrix} \tag{3}$$

for some $m > n$, and define

$$A = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ \rho_1 & \rho_2 & \cdots & \rho_n \\ \vdots & \vdots & & \vdots \\ \rho_1^{m-1} & \rho_2^{m-1} & \cdots & \rho_n^{m-1} \end{bmatrix} \tag{4}$$

$$x(t) = \begin{bmatrix} \alpha_1\rho_1^t \\ \\ \alpha_n\rho_n^t \end{bmatrix}; \; e(t) = \begin{bmatrix} \tilde{e}(t) \\ \vdots \\ \tilde{e}(t+m-1) \end{bmatrix} \tag{5}$$

With the above definitions we can write

$$\begin{aligned} y(t) &= \begin{bmatrix} 1 & 1 & \cdots & 1 \\ \rho_1 & \rho_2 & \cdots & \rho_n \\ \vdots & \vdots & & \vdots \\ \rho_1^{m-1} & \rho_2^{m-1} & \cdots & \rho_n^{m-1} \end{bmatrix} \begin{bmatrix} \alpha_1\rho_1^t \\ \\ \alpha_n\rho_n^t \end{bmatrix} + e(t) \\ &= Ax(t) + e(t). \end{aligned} \tag{6}$$

The key equation here is (6), which resembles the "standard" model used in sensor array signal processing. We form the following covariance matrix

$$\hat{R}_d = \sum_{k=1}^{M} y\left((k-1)d+1\right) y^*\left((k-1)d+1\right) \tag{7}$$

where $d \geq 0$ is an integer which controls the degree of overlap, between adjacent snapshots (the smaller the d the more overlapped those vectors are; for $d \geq m$ there is no overlapping), and M is the total number of snapshots defined as

$$M = \left\lfloor \frac{N-m}{d} \right\rfloor + 1$$

with $\lfloor\cdot\rfloor$ denoting rounding to the nearest smaller integer. The reason for the introduction of d is that when d is increased the noise vectors in (6) become less correlated (they are uncorrelated for $d \geq m$) which might improve the statistical performance in spite of the fact that M decreases as d increases.

For sufficiently large SNR values $\hat{R}_d$ in (7) is close to the matrix

$$R_d = AP_dA^* \tag{8}$$

where

$$P_d = \sum_{i=1}^{M} x\left((i-1)d+1\right) x^*\left((i-1)d+1\right). \tag{9}$$

Let

$$X \triangleq \begin{bmatrix} \alpha_1\rho_1 & & 0 \\ & \ddots & \\ 0 & & \alpha_n\rho_n \end{bmatrix} \begin{bmatrix} 1 & \rho_1^d & \cdots & \rho_1^{(M-1)d} \\ \vdots & \vdots & & \vdots \\ 1 & \rho_n^d & \cdots & \rho_n^{(M-1)d} \end{bmatrix}.$$

It can readily be verified that

$$P_d = XX^* \tag{10}$$

and consequently, from basic rank properties of Vandermonde matrices

$$\mathrm{rank}\,(P_d) = \min(M,n) \triangleq \bar{n}. \tag{11}$$

It seems reasonable to assume that $M \geq n$, and hence that $\bar{n} = n = \mathrm{rank}\,(R_d)$. This means that n can be obtained as the "practical" rank of $\hat{R}_d$, and the assumption that n is known can be relaxed in practice.

Suppose we know n. Let $\{b_k\}_{k=0}^n$ be the coefficients of the following polynomial

$$b_0z^n + \cdots + b_{n-1}z + b_n = b_0\prod_{k=1}^{n}(z-\rho_k)$$

and let

$$B^* = \begin{bmatrix} b_n & \cdots & b_0 & & 0 \\ & \ddots & & \ddots & \\ 0 & & b_n & \cdots & b_0 \end{bmatrix} \quad (m-n)\times m. \tag{12}$$

Also define the eigenvalue decomposition of R_d as

$$R_d = \begin{bmatrix} S & G \end{bmatrix} \begin{bmatrix} \Lambda & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} S^* \\ G^* \end{bmatrix} = S\Lambda S^* \quad (13)$$

where S is the matrix whose columns are the n principal eigenvectors of R_d and Λ is a diagonal matrix with the corresponding eigenvalues on the diagonal. It is well known that $\mathcal{R}(S) = \mathcal{R}(A)$, where $\mathcal{R}$ denotes the range-operator. Consequently, as $B^*A = 0$ (as is readily verified) we have

$$B^*S = 0. \quad (14)$$

This is a key property whose potential for parameter estimation is rather obvious. Let $\hat{S}$ and $\hat{\Lambda}$ denote the sample counterparts of S and Λ as defined in (13). In view of (14), we can expect that the following equation in the unknowns $\{b_k\}_{k=0}^n$ (which are used to reparameterize the estimation problem under discussion) holds approximately (for SNR $\gg 1$):

$$B^*\hat{S} \simeq 0. \quad (15)$$

Equation (15) can be implemented by minimizing the quadratic cost function

$$\begin{aligned} f(b) &= \mathrm{tr}\left[(W_1^{1/2}B^*\hat{S}W_2^{1/2})(W_1^{1/2}B^*\hat{S}W_2^{1/2})^*\right] \\ &= \mathrm{tr}\left[BW_1B^*\hat{S}W_2\hat{S}^*\right]. \end{aligned} \quad (16)$$

We show in the Appendix that good choices of the weighting matrices are $W_2 = \Lambda$ and $W_1 = (B^*B)^{-1}$ (or consistent estimates of these quantities). A more intuitive argument for these choices is:

1. Since the eigenvectors in $\hat{S}$ are determined with an accuracy that is proportional to the square root of the corresponding eigenvalues, it makes sense to post-weight the equations in (15) via $\hat{\Lambda}^{1/2}$.
2. Experience with similar problems in array signal processing and system identification [2, 9, 10, 11] suggests that improved numerical and statistical accuracy is obtained if row weighting is used to make B column unitary, which is achieved with $W_1 = (B^*B)^{-1}$.

We note that, in view of (14), minimization of (16) with W_1 replaced by any positive definite matrix (such as I) gives *consistent (in SNR) estimates* of $\{b_k\}_{k=0}^n$. Furthermore, it can be shown that, asymptotically in SNR, replacement of the weight $W_1 = (B^*B)^{-1}$ in (16) by a consistent estimate has *no effect on the asymptotic accuracy.* Hence the following two-step procedure appears suitable to use for minimization of (16).

Step 1. Compute the n principal eigenpairs of $\hat{R}_d$. Let

$$f_W(b) = \mathrm{tr}\left[BW^{-1}B^*\hat{S}\hat{\Lambda}\hat{S}^*\right]. \quad (17)$$

Obtain initial (consistent) estimates of b by minimizing $f_W(b)$, with $W = I$.

Step 2. Derive enhanced estimates of $\{b_k\}$ as the minimizer of $f_W(b)$, with $W = \hat{B}^*\hat{B}$ and where $\hat{B}$ is made from the estimates obtained in Step 1. Obtain $\{\rho_k\}$ from $\{b_k\}$.

Both the minimization steps above can be efficiently performed by the algorithm outlined in the following subsection.

3.1 Minimizing $f_W(b)$

We start out from the following form of the cost function in (16):

$$\begin{aligned} f(b) &= \mathrm{vec}(B)^*\left(W_1^T \otimes \hat{S}W_2\hat{S}^*\right)\mathrm{vec}(B) \quad (18) \\ &= \tilde{b}^*\left(W_1^T \otimes \hat{S}W_2\hat{S}^*\right)\tilde{b}. \quad (19) \end{aligned}$$

where vec denotes the vectorization operator, $\otimes$ denotes the Kronecker product and

$$\begin{aligned} \tilde{b}^* &= \mathrm{vec}(B)^* = \begin{bmatrix} b^T & 0^T & b^T & 0^T & \cdots \end{bmatrix} \quad (20) \\ b^T &= \begin{bmatrix} b_n & \cdots & b_0 \end{bmatrix}. \quad (21) \end{aligned}$$

If we use $\bar{\Omega}$ to denote the matrix $W_1^T \otimes \hat{S}W_2\hat{S}^*$ from which the *rows and columns corresponding to the zeros in $\tilde{b}$ are eliminated,* and also denote by Ω the following matrix

$$\Omega^T = \begin{bmatrix} I & \cdots & I \end{bmatrix} \bar{\Omega} \begin{bmatrix} I \\ \vdots \\ I \end{bmatrix}, \quad (22)$$

then (19) can be written as

$$f(b) = b^*\Omega b. \quad (23)$$

The function (23) is to be minimized with respect to b, under an appropriate constraint. If we choose a unit norm constraint on b we get the total least squares solution (TLS) which is easily obtained as the eigenvector of Ω corresponding to the smallest eigenvalue. In summary

$$\boxed{\hat{b}\text{=the smallest eigenvector of } \Omega.}$$

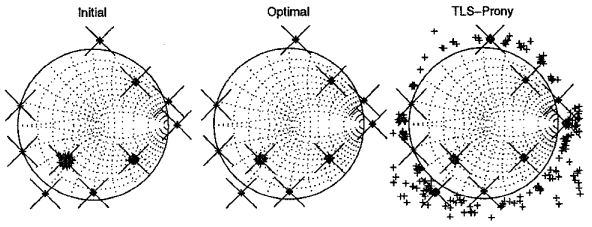

Figure 1: Locations of 100 pole estimates calculated with the proposed method and the TLS-Prony method. The leftmost graph corresponds to Step 1 in the MODE algorithm and the middle one to Step 2.

4 Numerical examples

4.1 Example 1

Figure 1 compares the proposed method with the TLS-Prony [8] method. The example is adopted from [8]. There are ten exponential modes, which are selected in such a way that the scenario is a rather general one. The true pole locations are indicated with large '×'s in Figure 1. There were $N = 100$ data points and $\sigma = 0.01$. The amplitude coefficients, $\{\alpha_k\}_{k=1}^{10}$, are chosen so that each mode energy is unity. This corresponds to an SNR of 20 dB per mode. In the MODE algorithm $d = 1$ and $m = 20$ are used. In Figure 1 the '+' signs show the pole-estimates obtained from 100 independent Monte-Carlo simulations with the proposed algorithm (both initial and optimal estimates) and the TLS-Prony algorithm discussed in [8]. The proposed algorithm produces more reliable estimates and in addition it is computationally simpler and more straightforward to implement.

4.2 Example 2

The next example investigates, empirically, the performance of the MODE algorithm as a function of d and m. The scenario is the same as in Example 1. In Figure 2 the sum of the RMSE's of the ten modes, calculated from 100 independent realizations, is displayed as a function of d and m. The user's parameters d and m are varied from 1 to 25 and from 10 to 30, respectively. The result in Figure 2 suggests that d should be chosen small and that m should be chosen as approximately twice the number of modes, for this particular scenario.

A Optimal Weight Selection

For two general weights W_1 and W_2 we can write the cost-function (*c.f.* (16))

$$f(b) = \mathrm{vec}\,(B)^* \left(W_1^T \otimes \hat{S} W_2 \hat{S}^*\right) \mathrm{vec}\,(B) \qquad (24)$$

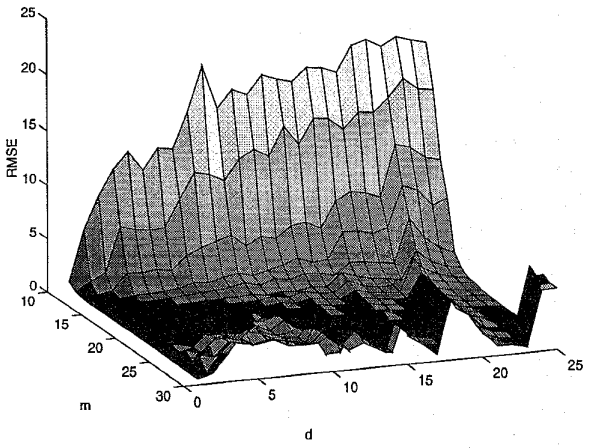

Figure 2: Sum of the RMSE's as a function of d and m

Since $\hat{b}$ is a consistent (in SNR) estimate of b we have

$$\hat{b} - b \simeq -f''(b)^{-1} f'(b), \qquad (25)$$

which can be simplified, by using calculations similar to what is done in [2]:

$$\hat{b} - b \simeq (D^* Q D)^{-1} D^* Q \,\mathrm{vec}\left(\hat{S}^* B\right) \qquad (26)$$

$$D = \left[\left(I \otimes \hat{S}^*\right) \frac{\partial \tilde{b}}{\partial b_1} \cdots \left(I \otimes \hat{S}^*\right) \frac{\partial \tilde{b}}{\partial b_n}\right] \qquad (27)$$

where $Q = W_1^T \otimes W_2$ and $\tilde{b} = \mathrm{vec}(B)$. Hence in order to calculate the covariance matrix of $\hat{b} - b$

$$\mathrm{E}\left(\hat{b} - b\right)\left(\hat{b} - b\right)^* = (D^* Q D)^{-1} D^* Q \left[\mathrm{E}\,\mathrm{vec}\left(\hat{S}^* B\right) \mathrm{vec}\left(\hat{S}^* B\right)^*\right] Q D \,(D^* Q D)^{-1}, \qquad (28)$$

we need to calculate the high SNR covariance matrix of $\mathrm{vec}(\hat{S}^* B)$. A rather lengthy asymptotic analysis, included in the full version of the paper, shows that

$$\mathrm{E}\left[\mathrm{vec}\left(\hat{S}^* B\right) \mathrm{vec}\left(\hat{S}^* B\right)^*\right] \simeq \simeq \sum_{\xi=-M}^{M} (M - |\xi|) \left[B^T C_d^T(\xi) \bar{B}\right] \otimes \left[\Lambda^{-1} S^* A P_d(\xi) A^* S \Lambda^{-1}\right]. \qquad (29)$$

where the overbar denotes comples conjugate,

$$C_d(\xi) = \mathrm{E}\,(e(t)\, e^*(t + \xi d)) \qquad (30)$$

is the noise covariance function and

$$P_d(\xi) = \sum_{i=-\min(0,\xi)}^{M-\max(0,\xi)-1} x(id+1)\, x^*((i+\xi)d+1). \qquad (31)$$

The expression (29) for the covariance of $\mathrm{vec}(\hat{S}B)$ is in general not in a Kronecker product form. This implies that the trace-form criterion (16) is not general enough to get optimal accuracy for the estimates of b. However, previous experience (see *e.g.* [2]) suggests that satisfactory accuracy can be obtained by truncating the sum in (29) to include only the term with $\xi = 0$, *i.e.*

$$\mathrm{E}\left(\hat{b}-b\right)\left(\hat{b}-b\right)^{*} \approx M\left(D^{*}QD\right)^{-1} D^{*}Q \left[\left(B^{T}C_{d}^{T}(0)\bar{B}\right)\otimes\Lambda^{-1}\right] QD\left(D^{*}QD\right)^{-1} \quad (32)$$

where we have used the fact that

$$S^{*}\left[\sum_{i=1}^{M}\left(Ax\left((i-1)d+1\right)x^{*}\left((i-1)d+1\right)A^{*}\right)\right]S = \Lambda.$$

The expression (32) is minimized by choosing the weighting matrix Q as

$$Q = \left(W_1^{T}\otimes W_2\right) = \left[\left(B^{T}C_{d}^{T}(0)\bar{B}\right)\otimes\Lambda^{-1}\right]^{-1}$$

i.e. $W_1 = \left(B^{*}C_d(0)B\right)^{-1}$ and $W_2 = \Lambda$. The above calculations justifies the weights chosen in Section 3, under the white noise assumption (white noise $\Rightarrow$ $C_d(0) = \sigma^2 I$).

References

[1] R. Carrière and R. L. Moses. High resolution radar target modeling using a modified Prony estimator. *IEEE Transactions on Antennas and Propagation*, 40(1):13–18, Jan. 1992.

[2] M. Cedervall and P. Stoica. System identification from noisy measurements by using instrumental variables and subspace fitting. *Circuits, Systems and Signal Processing*, 1995. To appear.

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

[4] B. D. Jaffe, S. W. Chen, L. C. Potter, and S. D. Nelson. Prony frequency analysis of prolonged qrs signal averaged ecgs. In *Circ. Research*, volume 41, page 659A, 1993. (Presented to the Midwest Meeting of the American Federation of Clinical Research, Chicago, IL, November 1993).

[5] A. K. Krishnamurthy. Glottal source estimation using a sum-of-exponentials model. *IEEE Transactions on Signal Processing*, 40(3):682–686, Mar. 1992.

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

[7] B. D. Rao and K. S. Arun. Model Based Processing of Signals: A State Space Approach. *Proceedings of the IEEE*, 80:283–309, Feb. 1992.

[8] W. M. Steedly, C. J. Ying, and R. L. Moses. Statistical analysis of TLS-based Prony techniques. *Automatica, Special Issue on Statistical Signal Processing and Control*, 30(1):115–129, Jan. 1994.

[9] P. Stoica and A. Nehorai. MUSIC, maximum likelihood, and Cramér-Rao bound. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, ASSP-37(5):720–741, May 1989.

[10] P. Stoica and A. Nehorai. Performance study of conditional and unconditional direction-of-arrival estimation. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, ASSP-38:1783–1795, October 1990.

[11] P. Stoica and K. C. Sharman. Maximum likelihood methods for direction-of-arrival estimation. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, ASSP-38(7):1132–1143, July 1990.