A Combined Order Selection and Parameter Estimation Algorithm for Undamped Exponentials

Ching-Hui J. Ying, Ashutosh Sabharwal, and Randolph L. Moses, *Senior Member, IEEE*

***Abstract*—We propose an approximate maximum likelihood parameter estimation algorithm, combined with a model order estimator, for superimposed undamped exponentials in noise. The algorithm combines the robustness of Fourier-based estimators and the high-resolution capabilities of parametric methods. We use a combination of a Wald statistic and a MAP test for order selection and initialize an iterative maximum likelihood descent algorithm recursively based on estimates at higher candidate model orders. Experiments using simulated data and synthetic radar data demonstrate improved performance over MDL, MAP, and AIC in cases of practical interest.**

***Index Terms*—Combined detection and estimation, resolution bounds, undamped exponentials.**

I. Introduction

THIS PAPER presents an algorithm for the simultaneous estimation of parameters and detection of model order for a noisy exponential sequence. It represents an attempt to find a parameter estimation algorithm that is capable of high resolution (resolving exponential components separated by less than the Fourier resolution limit) that provides accurate modeling of measured signals with practical model orders (10–50), that is applicable for practical data lengths and signal-to-noise ratios, that does not require much tuning by the user, and that fails gracefully. We attempt to combine the robustness of nonparameteric estimators with the increased resolution of parametric estimators and combine order selection with parameter estimation.

Fourier-based estimation methods faithfully model the highest energy components in the signal. They are robust in the sense that high energy components are always modeled. The resolution limit of the periodogram hinders its ability to resolve closely spaced terms, and parameter estimation bias sometimes results from this resolution limit. Modified periodogram techniques, including CLEAN and its variants [1], [2] and notch-periodogram methods [3], [4] are somewhat helpful in reducing bias but still may fail to reliably resolve closely spaced modes and exhibit significant bias unless the modes are well separated.

Subspace-based methods, including MUSIC, ESPRIT, Min-Norm, and matrix pencil methods, provide increased resolution over Fourier-based methods; however, they can miss dominant peaks in the signal. For example, a subspace method can fail to model high-amplitude signal components or can fail to resolve modes that are close but still separated above the Fourier resolution limit. This behavior stems from the nonlinear effect of separating a signal subspace from a noise subspace; unfortunately, it is difficult to predict how the subspace separation will affect modeled energy in signals and, therefore, difficult to design around subspace separation failures.

Iterative methods, including ML-based methods [5]–[7], can give good results if the iterative descent procedures are well initialized. Unfortunately, good initialization is difficult because the likelihood function being maximized is highly multimodal [6]. Initialization by either a Fourier-based or a subspace-based method often results in convergence to a local minimum of the loss function, resulting in a suboptimal modeling accuracy. The only reliable way to ensure convergence to a global optimum is to use an expensive grid search for initialization; the result is an algorithm that is often computationally intractable in practice.

Turning to order detection, there are a number of techniques available, including AIC [8], MDL [9], MAP techniques [10], [11], and others [12], [13]. Many of these methods can be shown to have nice asymptotic properties (e.g., consistent estimation of model order as either data length or SNR increases). However, these techniques may be slow to converge to their asymptotic performance and may perform poorly for cases of small data length or finite SNR that occur in practice. In addition, most methods require obtaining parameter estimates for every model order from one to some maximum order and do not use any model structure to reduce computation burden.

Combined detection/estimation methods have also been considered; see, e.g., [3], [4], [11], [14], and [15]. The methods in [3] and [4] are based on periodogram analysis along with spectral notching and provide computationally fast but statistical suboptimal solutions. Methods in [11], [14], and [15] are statistically optimal or close to optimal but require searching a large number (usually all possible) model orders. The goal in this paper is to provide a statistically (near) optimal method that has a moderate computational cost. Computation is saved by i) estimating parameters on only a subset of possible model orders and ii) initializing iterative descent procedures intelligently based on estimates obtained from previous searches.

In this paper, we present a hybrid nonparameteric–parametric method for combined order selection and maximum likelihood

Manuscript received March 19, 1998; revised February 15, 1999. This work was supported in part by the Defense Advanced Research Project Agency under Contract MDA972-93-1-0015 managed through Air Force Research Laboratories. The associate editor coordinating the review of this paper and approving it for publication was Prof. James A. Bucklew.

C.-H. J. Ying was with the Department of Electrical Engineering, The Ohio State University, Columbus, OH 43210-1272 USA. He is now a Buddhist Monk in Taiwan.

A. Sabharwal was with the Department of Electrical Engineering, The Ohio State University, Columbus, OH 43210-1272 USA. He is now with Rice University, Houston, TX 77005 USA (e-mail: ashu@rice.edu).

R. L. Moses is with the Department of Electrical Engineering, The Ohio State University, Columbus, OH 43210-1272 USA (e-mail: randy@ee.eng.ohio-state.edu).

Publisher Item Identifier S 1053-587X(00)01538-5.

parameter estimation. The algorithm initializes using a nonparameteric estimate (such as a periodogram estimate). We choose the periodogram because it is computationally efficient, as well as nearly statistically efficient if the exponential modes are well separated. When the exponential modes are not well separated, the periodogram fails to resolve them but still provides information useful for initialization of an iterative descent parametric estimator. We split initial periodogram peaks to initialize a high-resolution parametric estimator and recursively decrease a resultant high model order to obtain a final model order estimate and corresponding parameter estimates. Decisions on reducing model order are based on a Wald statistic coupled with a maximum *a posteriori* (MAP) order selection rule. Robustness and graceful failure of the estimation procedure are achieved by periodogram-based initialization; we start with the periodogram as a baseline performance and improve on it but avoid hard failures (such as failing to estimate a strong modal component) associated with parametric techniques. The increased resolution of parametric modeling provide a method to improve on the periodogram estimates when there is sufficient SNR to do so. In addition, the algorithm exploits the nested structure of the candidate models to obtain a joint detection-estimation method where estimates from one model order affect the subsequent estimation as the model order is updated.

The paper is organized as follows. In Section II, we present the proposed order selecting maximum likelihood (OSML) algorithm, and we discuss some of the practical issues related to algorithm implementation. Section III presents some practical numerical examples that illustrate the effectiveness of the algorithm. Finally, Section IV provides some conclusions and related observations.

II. Combined Detection/Estimation Algorithm

The noise perturbed undamped exponential model is given by

$$y(t) = s(t) + \eta(t) = \sum_{i=1}^{q} \alpha_i e^{\jmath\phi_i} e^{\jmath\omega_i t} + \eta(t) \qquad t = 1, \ldots, N \quad (1)$$

The model order q and the model parameters $\{\alpha_i, \phi_i, \omega_i\}_{i=1}^{q}$ are unknown and need to be estimated from the observations $\{y(t)\}_{t=1}^{N}$. The noise $\eta(t)$ is assumed to be independent of $s(t)$, i.i.d., and Gaussian with mean zero and variance σ^2. We also assume that an upper bound q_u on the model order is available.

The proposed algorithm estimates q and $\{\alpha_i, \phi_i, \omega_i\}_{i=1}^{q}$ and can be divided into the following three steps.

1) *Fourier-Based Initialization*: The initialization step selects dominant peaks using a Fourier-based method (such as the periodogram) and splits them into a several equal energy modes as a first step in achieving higher resolution. The amount of splitting is determined by the estimated CRB for the frequency parameter.
2) *ML Refinement*: A gradient descent of the likelihood function is performed, using a fixed model order and the initialization as described in Step 1 above.
3) *Mode Combining*: The ML estimate is used to estimate the CRB, which in turn is used to combine extra modes. The combining algorithm performs a hypothesis test using the Wald statistic [16]. The ML refinement and mode combining steps continue until the order selection criterion indicates that no more modes are to be combined. Finally, the algorithm uses MAP [11] to reject low energy modes.

A. Step I: Fourier-Based Initialization

The basic idea of the Fourier-based initialization is as follows. Dominant peaks from the periodogram of the $\{y(t)\}$ are chosen to obtain an initial estimate of model order, model parameters, and noise variance. Based on the estimated SNR for each mode, a best-case CRB is computed to obtain the resolution limits of multiple closely spaced modes that might give rise to this periodogram peak. Each periodogram peak is split into several modes, where the number and the placement of the modes is determined by the estimated CRB.

Dominant peaks are found by selecting peaks in a periodogram spectral estimate that are above $\hat{\sigma}^2$, which is an estimate of σ^2. A consistent estimate of σ^2 is given as [17]

$$\hat{\sigma}^2 = \frac{1}{M - q_u} \sum_{k=q_u+1}^{M} \lambda_k \quad (2)$$

where $\{\lambda_k\}_{k=1}^{M}$ are the eigenvalues of matrix $(1/(N - M + 1))Y^*Y$, with

$$Y = \begin{bmatrix} y_1 & y_2 & y_3 & \cdots & y_M \\ y_2 & y_3 & y_4 & \cdots & y_{M+1} \\ y_3 & y_4 & y_5 & \cdots & y_{M+2} \\ \vdots & \vdots & \vdots & & \vdots \\ y_{N-M} & y_{N-M+1} & y_{N-M+2} & \cdots & y_N \end{bmatrix} \qquad M > q_u, \quad N - M > q_u \quad (3)$$

and q_u is an upper bound of the model order. The advantage of using this technique is that no information about the true signal (which we want to estimate) is needed; the only information needed is an upper bound of the model order. Fast subspace decomposition [18] can be used to considerably ease the computational burden of the eigendecomposition. The estimate of σ^2 is used to threshold the periodogram. Values in the periodogram above $\kappa\hat{\sigma}^2$ are retained as dominant peaks (we use $\kappa = 3$ in our simulations).

Each detected peak is split into a number of "super-resolved" frequency estimates (i.e., separated by less than the Fourier resolution limit) using an estimate of the best case CRB to determine the frequency separation. For each mode, the SNR is estimated as

$$\widehat{\text{SNR}}/\text{mode} = \frac{\hat{\alpha}_i^2}{\hat{\sigma}^2}. \quad (4)$$

Having selected the dominant peaks of the periodogram, each of the peak is split into v modes, as described below.

We assume the peak results from two equal amplitude modes (with each amplitude $\hat{\alpha}/\sqrt{2}$) that are closely spaced in frequency and ask what is the minimum frequency separation for the two modes to be resolvable using the CRB as the bound on standard deviation of frequency estimates. Consider two poles on the unit circle $p_1 = e^{\jmath\omega}$ and $p_2 = e^{\jmath(\omega+\Delta\omega)}e^{\jmath\phi}$, where $\Delta\omega$ is the frequency separation. To define a detection resolution limit for closely spaced modes, we note that the resolvability

of two equal energy modes depends on the modes only through their separation $\Delta\omega$. We propose a resolution limit $u_\phi(\Delta\omega)$[1]

$$u_\phi(\Delta\omega) \triangleq \sigma\sqrt{\left(\sigma_\omega^2 + \sigma_{\omega+\Delta\omega}^2 - 2\sigma_{\omega,\omega+\Delta\omega}\right)} \tag{5}$$

where

$$\mathrm{CRB}_\phi(\omega, \omega+\Delta\omega) = \sigma^2 \begin{bmatrix} \sigma_\omega^2 & \sigma_{\omega,\omega+\Delta\omega} \\ \sigma_{\omega,\omega+\Delta\omega} & \sigma_{\omega+\Delta\omega}^2 \end{bmatrix}$$

is the CRB matrix of the angles of p_1 and p_2, respectively. The function $u_\phi(\Delta\omega)$ is interpreted as the minimum value of frequency separation for which two equal energy amplitude modes can be resolved. Expressions for the CRB can be found in [19].

The resolution limit $u_\phi(\Delta\omega)$ also admits a geometric interpretation. The Fisher information matrix is a Riemannian metric on the statistical manifold obtained by varying the parameters $\{\alpha_i, \phi_i, \omega_i\}$ [20]. In other words, it is a local measure of distance on the manifold of possible true models. The quantity $(\Delta\omega)^2/[u_\phi(\Delta\omega)]^2$ is proportional to the distance (induced by the Riemannian metric) of $\Delta\omega$ from $\Delta\omega = 0$ at high SNR. We believe that this measure of distance is the natural quantity to define the frequency resolution bound; this idea is further elaborated in [21]. Detection thresholds for closely spaced modes have been defined and computed in [22]–[24]; the previous work has focused only on the diagonal entries of CRB. For closely spaced modes ($\Delta\omega < (2\pi/N)$), the off-diagonal entries are of the same order (irrespective of the data length) as the diagonal entries and, hence, cannot be ignored; the proposed resolution bound (5) uses the complete CRB. For well-separated modes, only diagonal entries of CRB are dominant and, hence, the proposed limit reduces to the previously proposed thresholds.

Note that $u_\phi(\Delta\omega)$ in (5) also gives a detection lower bound for all combinations of signals (e.g., two unequal energy modes, more than two modes, etc.); for two unequal energy modes, the smaller energy mode will have a larger CRB than the CRB of the higher energy mode, and consequently, the resolution limit is larger than the one for the equal energy mode case (if choosing the energy in the equal energy case to match the energy of the higher energy mode). In addition, the two-mode resolution limit is a lower bound for all multiple mode cases; this is so because if three or more modes are closely spaced, the CRB's of these modes are higher than the CRB's for the two-mode case (if choosing the shortest distance of the multiple mode case as the distance of the two-mode case) [25] and, consequently, give a higher resolution limit.

The resolution limit is found by using (5). We set $u_\phi(\Delta\omega) = \gamma\Delta\omega$ for some user-selected choice of γ and determine the set of $\Delta\omega$ values that solve (5). For high SNR, $\gamma u_\phi(\Delta\omega)$ defines the distance between two modes with which γ standard deviation CRB confidence intervals of the frequency estimates of the two modes become disjoint; in particular

$$\Pr(|\Delta\omega| < \gamma u_\phi \mid \mathrm{SNR}, N) > (1-p) = \Phi(\gamma)$$

where $\Phi(x) = (2/\sqrt{2\pi})\int_0^x e^{-(x^2/2)}dx$ is the error function. In this paper, we use $\gamma = 2$, which gives a 95% CRB confidence interval. The effect of γ on over/under-modeling errors is discussed in Section II-D.

[1]The subscript ϕ is used to emphasize the dependence of resolution limit on the initial phase difference.

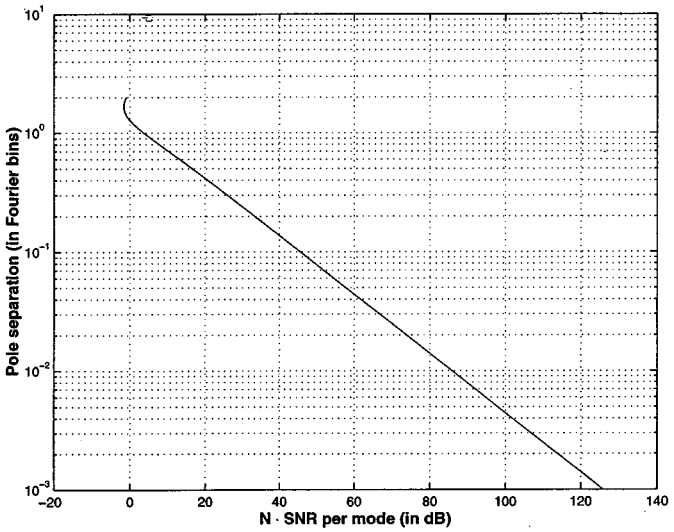


Fig. 1. Best-case two-mode resolution bound as a function of SNR.

The resolution limit $u_\phi(\Delta\omega)$ depends on the initial phase difference between the two modes. To ensure that the initial model order is overestimated, we choose ϕ such that the resulting detection threshold is minimized, i.e., $u(\Delta\omega) = \min_\phi u_\phi(\Delta\omega)$. The best case phase difference occurs at $\phi = ((m-1)\Delta\omega)/2$, i.e., the phase difference is zero at the center of the data record; this resolution limit is the smallest frequency resolution that can be supported by the data record. A plot of $(N\cdot\Delta\omega)/2\pi$ versus N·SNR is shown in Fig. 1. The normalization of $\Delta\omega$ by $2\pi/N$ gives the number of resolvable modes within a Fourier resolution bin (Fbin, i.e., $2\pi/N$ rad). The independence on data length of this curve (with appropriately normalized axes), and the linear trend can be derived analytically following an extension to [24] (the off-diagonal entries of the 2×2 CRB matrix are neglected in [24]). Given $\hat{u}$, the maximum number of resolvable modes within a Fourier bin is inversely proportional to the resolution limit and is given by

$$\hat{v} = \left\lceil \frac{2\pi}{N\cdot\hat{u}} \right\rceil \tag{6}$$

where $\lceil a \rceil$ denotes the smallest integer that is larger than a.

Initializing from an overparameterized model serves two purposes. First, it ensures that consistent estimates can be obtained (in the sense described in [21]). Second, this acts as a guard against an undermodeled initialization. Since the proposed procedure reduces a high model by recursively combining close modes, an undermodeled initialization leads to biased initial and, hence, biased final estimates.

Finally, the new peaks are obtained as follows. A peak in the kth Fbin is split into $\hat{v}_i$ modes placed at

$$\hat{\omega}_{k,i} = \begin{cases} \left(k \pm \frac{l+\frac{1}{2}}{\hat{v}_i}\right)\frac{2\pi}{N}, & l=1,\ldots,\frac{\hat{v}_i}{2}, \quad \text{if } \hat{v}_i \text{ is an even integer} \\ \left(k \pm \frac{l}{\hat{v}_i}\right)\frac{2\pi}{N}, & l=0,\ldots,\left\lfloor\frac{\hat{v}_i}{2}\right\rfloor, \quad \text{if } \hat{v}_i \text{ is an odd integer.} \end{cases} \tag{7}$$

We set the phases of the split peaks to that of the original peak. Each of the original peaks with magnitude $\hat{\alpha}_i$ is split into $\hat{v}_i$ peaks with magnitudes $\hat{\alpha}_i/\sqrt{\hat{v}_i}$.

B. Step II: ML Refinement

The mode estimates obtained from the initialization in Step I are refined using a gradient descent method. Under the given assumptions, the ML problem is the same as minimizing the following simplified negative log-likelihood function

$$L(\theta) = \sum_{t=0}^{N-1} |y_t - s_t(\theta)|^2. \tag{8}$$

Consequently, many existing iterative ML algorithms [5], [26]–[28] can be used to perform the required ML refinement procedure. A summary of several useful ML algorithms is provided in [29].

C. Step III: Mode Combining

By using a lower bound to compute the resolution limit, the number of modes in the initialization is generally overestimated. We could use a standard order selection procedure (such as MDL [9] or MAP [11]) at this point, using the initial number of modes as an upper bound on model order, but that entails testing all model orders. Instead, we recursively decrease model order with a sequential hypothesis test, and as we reduce model order, the parameter estimates from the higher order model provide a good initial estimate to the lower order log-likelihood descent computation.

The model order is reduced by a recursive binary hypothesis test. The hypothesis test is performed using the estimated CRB (at the ML estimates after Step II) followed by a test based on MAP [11] criterion. In combining two modes, the following binary hypothesis test is performed:

$$\begin{aligned} H_0 &: \text{Model order} \leq k-1 \\ H_1 &: \text{Model order} = k. \end{aligned} \tag{9}$$

In terms of the parameters, the above test can be written as

$$\begin{aligned} H_0 &: \omega_i = \omega_j \quad \text{for some } i \neq j \quad \text{OR} \\ &\quad \alpha_l = 0 \quad \text{for some } l \\ H_1 &: \text{Otherwise.} \end{aligned} \tag{10}$$

That is, either a single mode is modeled by two or more frequencies near that mode, or small amplitude mode(s) are present.

The above formulation requires simultaneous testing on both the frequency and the amplitude parameter. Because of the Fourier-based initialization outlined above, the modes tend to lie close to each other in groups. Hence, the hypothesis test in (10) is approximated by

$$\begin{aligned} H_0 &: \omega_i = \omega_j \quad \text{for some } i \neq j \\ H_1 &: \text{Otherwise} \end{aligned} \tag{11}$$

followed by

$$\begin{aligned} H_0 &: \alpha_l = 0 \quad \text{for some } l \in \{1, \ldots, k\} \\ H_1 &: \alpha_l \neq 0 \quad \forall l \end{aligned} \tag{12}$$

We first consider test (11). We seek to use the ML estimates of parameters for model order k to estimate the parameters for the model order $k-1$. The motivation of recursive estimation is as follows. First, a recursive method that utilizes estimates from the higher model order will tend to inherit the robustness of periodogram. Second, since the model form under consideration leads to nested model classes corresponding to different model orders, a recursive procedure appears to be most natural.

To obtain an initialization for model order $k-1$, we combine two modes into one mode from the peak estimates for model order k. For combining, we propose to use the Wald test, which requires estimates from order k only. The Wald test for (11) is

$$(\omega_i - \omega_j)^2 \left(\frac{\hat{u}_\phi}{\mu}\right)^{-2} \underset{\text{combine}}{\lessgtr} 1 \tag{13}$$

where $\hat{u}_\phi$ is recomputed from (5) using the ML estimates from Step II. In this paper, we choose $\mu = 2$; a brief discussion on choice of μ is given in Section II-D. Finally, the test (12) is performed using the MAP rule. The two steps are summarized below.

1) *Test whether two estimated modes correspond to a single true mode*. The combining is implemented as follows.
   a) Calculate the CRB at the ML estimates. This is done by inserting the ML parameter estimates and the noise power estimate $\hat{\sigma}^2$ [using (2)] into the CRB expression for the undamped exponential model [19]. Assuming that the current model order is $\hat{q}$, sort the frequency parameter vector $(\hat{\omega}_1, \hat{\omega}_2, \ldots, \hat{\omega}_{\hat{q}})$. Pick the modes with smallest frequency separation, and test for potential combining using the Wald statistic. If not combined, pick the pair of modes with the next smallest separation. If Wald's test suggests combining of two modes, go to Step 1b; else go to Step 2.
   b) Discard the lower energy mode from the pair of modes that fails the Wald test. Combine only one mode at a time, and repeat the ML refinement after each combining step.
2) *Test if any modes have small magnitudes*. In this case, we can use the MAP information criterion to eliminate low energy modes. Given a set of estimates and the corresponding MAP, we first discard a mode, which is the smallest energy mode. We then refine the estimates using the ML procedure and compute the corresponding MAP cost function. We compare the latter computed MAP cost MAP$(k-1)$ to the former cost MAP(k). If the MAP$(k-1)$ is smaller, we then use the second set of estimates as the new estimates. We repeat the procedure until the best model structure that has the smallest MAP cost is found. The MAP information criterion for our case is found to be

$$\text{MAP}(\hat{q}) = N \log(L(\hat{\theta})) + \frac{5\hat{q}}{2} \log(N) \tag{14}$$

where $L(\hat{\theta}), N$, and $\hat{q}$ are defined as above. The proposed algorithm is summarized in Table I.

The basic reason for separating the two tests is as follows. The Fourier-based initialization captures the dominant peaks in the data, which we then split into several closely spaced peaks to achieve high resolution. After the ML refinement, if two modes move closer, they will tend to have similar energies. The CRB being a lower bound on the estimation performance acts as a

TABLE I
SUMMARY OF THE OSML ALGORITHM

1. Obtain the estimated noise power $\hat{\sigma}^2$ using Equation (2). Using $3\hat{\sigma}^2$ as a threshold for the periodogram (computed using a Kaiser window with $\beta = 3$), obtain the locations, magnitudes, and phases of the periodogram peaks of the signal.
2. Calculate the CRB resolution limit, u_i (Equation (5)) and the maximum number of modes per Fbin, $\hat{v}_i$ (Equation (6)) for each peak. Split each peak location into $\hat{v}_i$ equal peaks as described in subsection II-A. Denote the total number of modes by $\hat{q}$.
3. Perform ML refinement on mode estimates using nonlinear minimization of the negative log-likelihood function (Equation (8)).
4. Compute the CRB at the ML estimate. Test whether two closely spaced modes should be combined using the Wald statistic (13).
5. If any of modes are discarded in step 4, $\hat{q} \leftarrow \hat{q} - 1$ and go to Step 3. If no mode is discarded, go to step 6.
6. Compute the MAP cost for $\hat{q}$ (Equation (14)). Discard the lowest energy mode, perform ML refinement, and compute MAP$(\hat{q} - 1)$.
7. If MAP$(\hat{q} - 1) <$ MAP$(\hat{q})$, then $\hat{q} \leftarrow \hat{q} - 1$, and go to step 6. Else, the estimation is complete.

certificate of faith in the resolved modes. If the estimated modes are closer than the CRB resolution limits, we choose to combine modes in accordance with the CRB.

Having tested on frequencies, low-energy modes are removed next. We have used a MAP test, but in fact, for the amplitude parameter (which is a linear parameter in additive Gaussian noise), other standard tests (MDL, Wald) give nearly identical detection results. In fact, for linear parameters in additive Gaussian noise, the Wald statistic can be obtained by scaling the generalized likelihood ratio statistic, leading to a test similar to MAP and MDL.

D. Choosing γ and μ

The choice of scalar parameters γ and μ denotes the tradeoff between the resolvability of close modes and correct detection of the model order. A small γ can be understood as a lesser faith in a Fourier-based method's resolvability capabilities and proceeds in a more conservative fashion by oversplitting. For example, $\gamma = 0$ implies splitting every mode into an infinite number of modes. On the other hand, a large value of μ tends to combine only very close modes, again proceeding cautiously. For $\mu = \infty$, we obtain (a slightly modified version of) MAP, which recursively computes ML estimates. Thus, a small value of the ratio γ/μ potentially leads to more overmodeling than undermodeling errors.

In our extensive simulations, we found that the algorithm's performance is very robust to the choice of μ for $\mu > 2$. We believe that the robustness comes from using Fourier-based initialization coupled with mode splitting. We support our claim with the following observation made during our extensive simulations. Since each periodogram peak is split into several closely spaced modes, the model order following frequency splitting usually upper bounds the true model order. Thus, each of the true modes is potentially modeled by more than one closely spaced mode. If any pair of the closely spaced modes is combined based on the frequency separation, the initialization guarantees (with high probability) that at least one of the modes is still close to the true mode. Since every combining is followed by an ML refinement, the new mode estimates tend to be closer to the true modes. As μ increases, the frequency combining leads to higher model order estimates; the bulk of model reduction is then performed by the MAP test. We propose $\gamma = 2$ and $\mu \in [2, 16]$ ($\mu = 2$ was used for simulations) as a possible set of parameters. Larger values of μ lead to high reliance on the MAP test, which in our simulations led to a loss in detection performance. A rigorous treatment on choice of the parameters remains a topic of future research.

E. Practical Issues

For long data length and/or high SNR cases, the CRB resolution limit usually will be small, and hence, the number of peaks resulting from the splitting algorithm will be large, and even unrealistic, for some cases. Although the procedure is needed for accurate detection of modes (since for large-sample and/or high SNR signals the signal frequencies can be very closely spaced), the price to pay is the high computational cost resulting from extremely large initial model orders. If it can be reasonably assumed in a given application that no frequencies are closer than u_{appl} rad, then the CRB resolution in (5) can be redefined as $u' = \max(u_{\text{appl}}, u)$.

III. SIMULATION STUDIES

In this section, we present three examples to demonstrate the performance of the proposed algorithm, which we refer to as the order-selecting maximum likelihood (OSML) algorithm. The first example illustrates the order-selecting ability and the high resolution of the OSML algorithm. For illustrative purpose, we detail all the steps of the algorithm using a single noise realization. In the second example, we choose a general ten mode radar-like signal. We generate 100 Monte Carlo simulations for different SNR's to collect performance statistics. Performance is evaluated in terms of order detection accuracy (via correct detection probability) and parameter estimation accuracy (via estimation variance). For order selection performance, we compare the OSML Monte Carlo results with the AIC, MDL, and MAP approaches. For estimation accuracy, we compare the OSML Monte Carlo results with the CRB results. To understand how the OSML algorithm performs on radar signals, in the third example, we use a synthetic radar backscatter signal and compare the OSML algorithm to the matrix pencil approach [30].

A. Example One: Closely Spaced Modes

In the first example, we choose a data sequence of length $N = 25$ and composed of four exponentials with parameters $\alpha_i = 1, \omega_i = \pm(2\pi/N), \pm 1.8(2\pi/N)$ and $\phi_i = 0, i = 1, \ldots, 4$. The additive Gaussian noise is set so that the SNR per mode is $10\log_{10}(\alpha_i^2/\sigma^2) = 0$ dB. This example considers two closely spaced frequencies that cannot be resolved by a Fourier-based

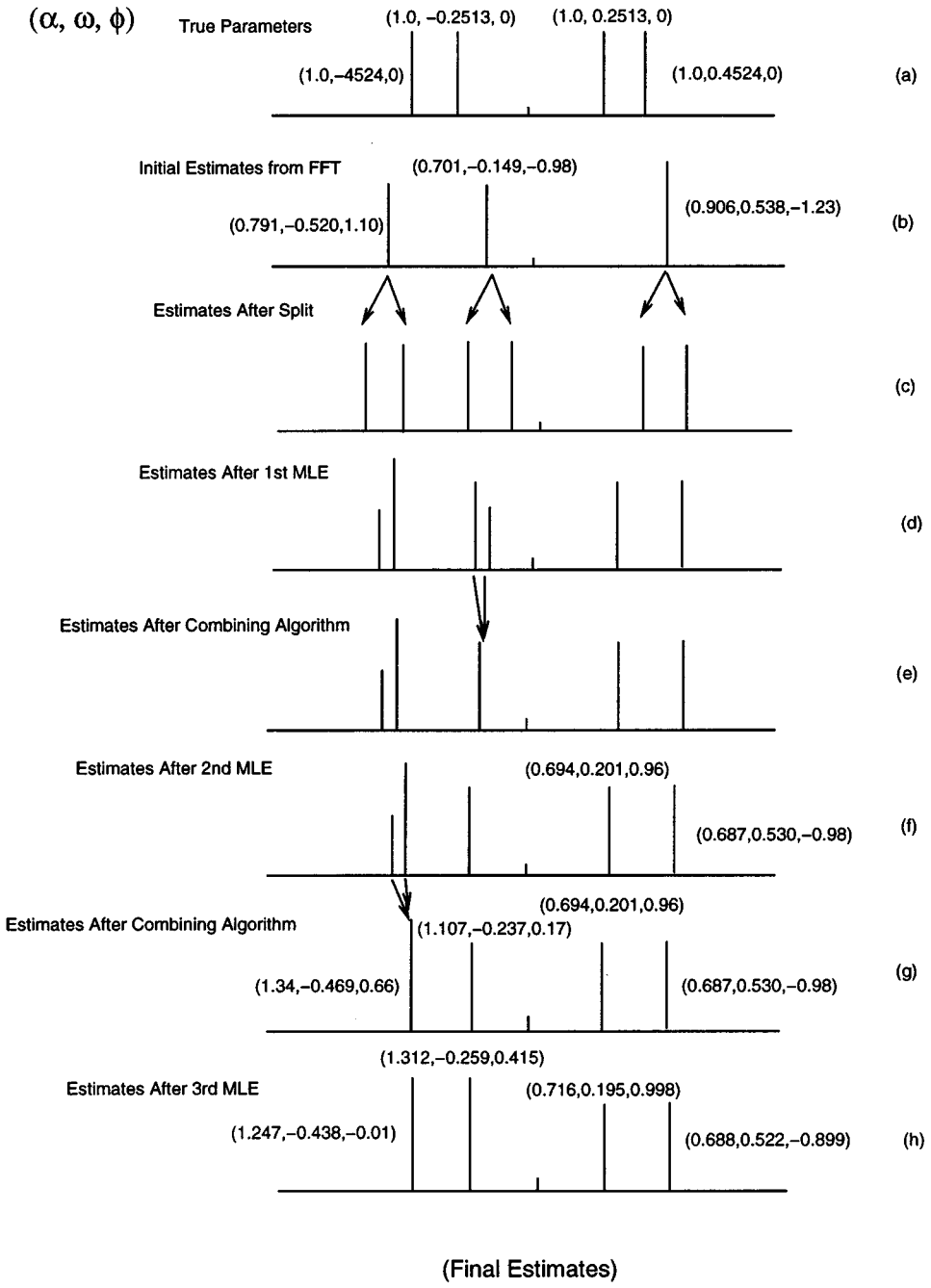


Fig. 2. Estimation results for Example One. The numbers in parentheses are the amplitudes, frequencies, and phases of the corresponding modes.

processing. This particular choice of parameters results in real-valued data; however, the estimation algorithm assumes complex-valued data.

Fig. 2 shows the entire estimation process for Example Two. The periodogram has only three peaks and does not resolve the closely spaced modes. From the CRB expressions, the maximum number of resolvable signals per Fbin is 2; therefore, for this case, there are six initial modes. The algorithm converges to the correct model order after three ML iteration steps, as shown in the figure. We see that the algorithm is able to determine the correct model order and resolve the closely spaced exponentials, even though the initial periodogram peaks did not resolve the modes.

B. Example Two: A General Ten Mode Case

We choose a ten-mode signal to demonstrate the statistical properties of the OSML algorithm. The spectrum of the true signal is shown in Fig. 3(a). In Fig. 3(b), we show the locations of the signal modes. All ten modes have equal energy, and the data length is $N = 64$. We run 100 Monte Carlo simulations for each SNR per mode value. We consider a moderate SNR range, where the SNR per mode ranges from -5 dB to 15 dB. We note that some of the modes are closely spaced (seen from Fig. 3(a)), and the periodogram cannot resolve all 10 modes and, hence, cannot be used to accurately estimate the model order.

In Fig. 4, we plot the probability of correct order selection versus SNR obtained with four order detection algorithms,

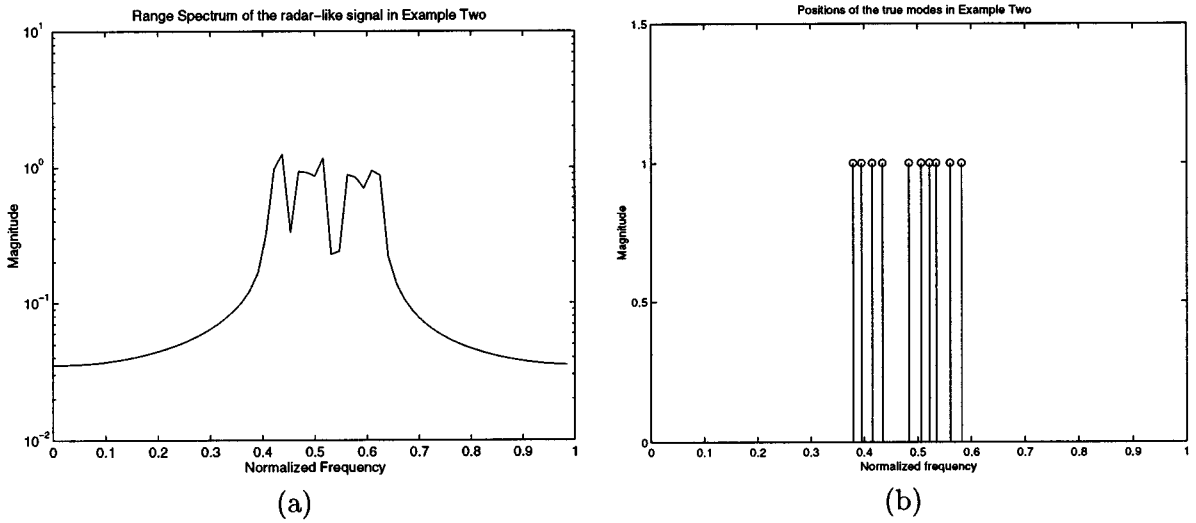


Fig. 3. (a) Spectrum and (b) mode locations of the radar-like signal in Example Two.

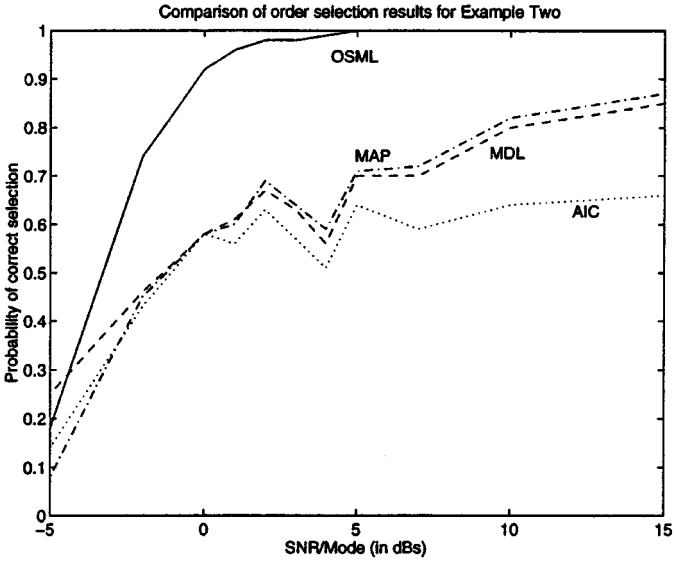


Fig. 4. Correct order selection probabilities of different order selection algorithms for Example Two.

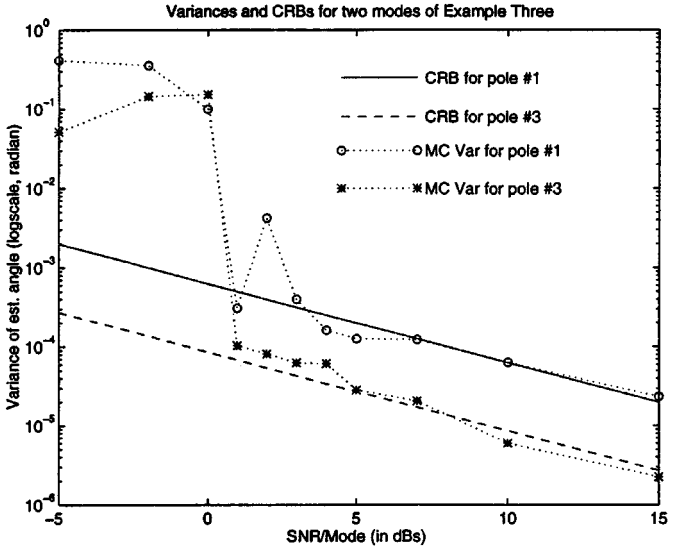


Fig. 5. Comparison of angle variances of Modes #1 and #3 for Example Two.

OSML, AIC [31], MDL [9], and MAP [11], [32]. The ML estimates for different orders used in each algorithm are the same and are obtained from the ML initialized with the OSML algorithm. The OSML algorithm stops when the order selection mechanism terminates; however, we calculate the ML estimates for all orders smaller than the upper bound for use by the other three order detection algorithms. The ML estimates are used in each of the criteria (AIC, MDL, and MAP), and the criteria are defined as

$$\hat{q}_{\mathrm{AIC}} = \arg\min_{q}\{m\log(L(\hat{\theta}_{\mathrm{ML}})) + 3q\} \tag{15}$$

$$\hat{q}_{\mathrm{MDL}} = \arg\min_{q}\left\{m\log(L(\hat{\theta}_{\mathrm{ML}})) + \frac{3q}{2}\log(m)\right\} \tag{16}$$

$$\hat{q}_{\mathrm{MAP}} = \arg\min_{q}\left\{m\log(L(\hat{\theta}_{\mathrm{ML}})) + \frac{5q}{2}\log(m)\right\}. \tag{17}$$

We see from the figure that the OSML algorithm performs substantially better than the other methods for this SNR range. The AIC, MDL, and MAP methods all tend to overestimate the model order (e.g., at SNR/Mode 10 dB of the figure, all mis-selections of the three algorithms are overestimation). For low model order signals (such as model order of 2), the performance of the other order selection algorithms are comparable to the OSML algorithm.

In Fig. 5, we compare the frequency estimation performance of the OSML algorithm with the corresponding CRB; we only show the performance for two of the ten modes (others are similar). We see that the OSML algorithm asymptotically achieves the CRB. The estimates appear to be slightly biased as the sample variances lie below the CRB. The bias can be partially explained by the finite grid of the Fourier-based initialization and the finite precision stopping rule of the ML optimization.

We also compared other parameter estimation algorithms, including the matrix pencil approach [30] and the IQML algorithm [5] for this particular set of data. The estimation variances are much worse because the modes are misestimated either due to high sensitivity of the subspace for matrix pencil

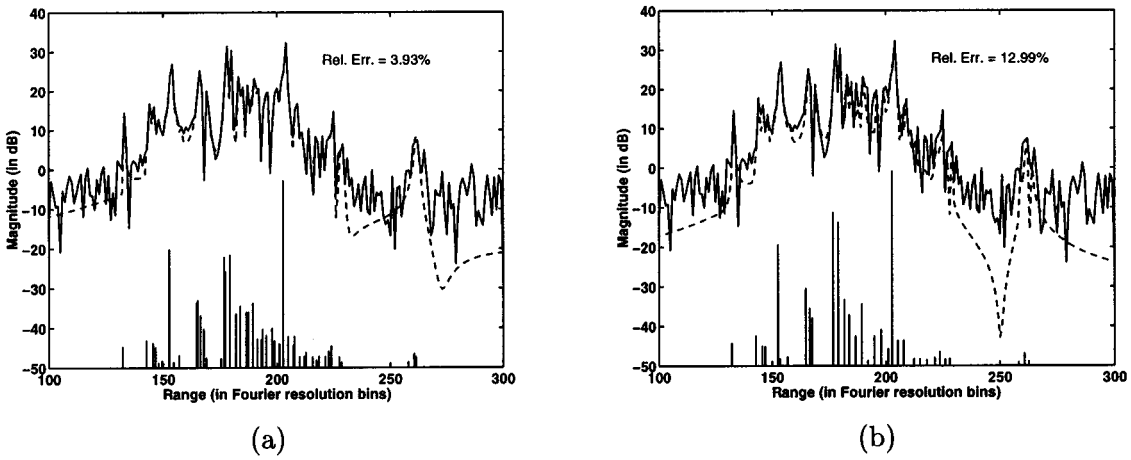


Fig. 6. Comparison of the OSML (a) and matrix pencil (b) methods for range profile modeling. XPatch generated range profile of an aircraft at X-band with SNR $= 13.89$ dB.

or poor initial estimates for IQML. The figure illustrates the robustness of the Fourier-based technique for use in practical applications and encourages the development of hybrid nonparameteric-parametric modeling techniques.

We note from Fig. 5 that the SNR threshold points for both modes are close to 0 dB; from Fig. 4, the SNR threshold point for good order detection is close to 2 dB. From the simulation results, it is evident that even though the model order is misestimated, the mode parameters, at least for some modes, are still accurately estimated using the OSML algorithm, which is a characteristic not found in the other parametric algorithms considered. In [22] and [23], it was shown that the detection threshold is lower than the estimation threshold for closely spaced modes. Our results show that even below the detection threshold, a subset of mode parameters can still be reliably estimated. In previous analysis [22], [23], a failure is declared if the model order is estimated incorrectly. Since, even with a detection error, reliable (unbiased) estimation of a part of the parameter set is still possible, the criteria for declaration of a failure may need to be modified.

C. Example Three: Simulated Radar Signal

The previous examples consider simulations using data that exactly match the assumed model. In Fig. 6, we show the estimation results for both the OSML algorithm and the matrix pencil algorithm applied to scattering center range estimates of X-band radar scattering of an aircraft synthesized using XPatch [33]. XPatch is used to find the frequency response of the aircraft from 9.5–10.5 GHz in 2.451 MHz steps. Although the XPatch data is not a sum of exponentials, it has been proposed to use exponential models to characterize high-frequency scattering data; the peaks in the transform domain (time or range in this case) correspond to scattering centers on the aircraft. The XPatch scattering measurements are corrupted by additive white Gaussian noise with SNR of 13.89 dB. For each algorithm, we show the original Kaiser-windowed periodogram range profile (solid lines) for comparison as ground truth and vertical lines to represent the locations of the estimated scattering centers. The reconstructed periodogram range profiles (broken lines) using the estimated scattering parameters at the original data bandwidth with periodogram resolution are also shown. The model order was autoselected to be 48 by the OSML method; for matrix pencil, both the AIC and MDL [34] predicted a model order of 33.

It is clear from the figure that the matrix pencil method fails to locate several of the higher energy scattering mechanisms, whereas the OSML method faithfully estimates all of the scattering centers that are within 25 dB of the peak. The matrix pencil algorithm, if used as an initial estimate to a ML descent procedure, results in convergence to a suboptimal local minimum because of the high initial bias. Finally, we note that the 3.93% residual error in the OSML estimate agrees very closely with the predicted 3.92% error that would result from perfect modeling of the target in white noise with SNR of 13.89 dB. The matrix pencil method with model order 48 also exhibits misestimation of dominant peaks, and the relative modeling error is 8.7%.

IV. Conclusion

In this paper, we presented a combined order selection and approximate maximum likelihood parameter estimation method for undamped exponential signals. The proposed approach addresses the key issue of reliable initialization for gradient descent methods to compute the ML estimates. The Fourier-based initialization results in an algorithm that models most of the signal energy, and the parametric estimation component provides superresolution when the SNR's of closely spaced modes are sufficiently high to support it. The algorithm depends on a few user-selectable parameters (i.e., resolution bounds, threshold in hypothesis testing), but algorithm performance seems to be fairly robust to selection of these parameters based on simulations performed. We demonstrate that order selection decisions using the Wald statistic [16] is effective for order estimation at noise levels and data lengths of practical interest.

Acknowledgment

The authors would like to thank the reviewers for helpful comments and suggestions, which helped streamline and focus

the presentation of the paper. They would also like to thank Prof. L. Potter of The Ohio State University for his insights on detection thresholds.

REFERENCES

[1] J. Tsao and B. D. Steinberg, "Reduction of sidelobe and speckle artifacts in microwave imaging: The CLEAN technique," *IEEE Trans. Antennas Propagat.*, vol. 36, pp. 543–556, Apr. 1988.

[2] J. Li and P. Stoica, "Efficient mixed-spectrum estimation with applications to target feature extraction," *IEEE Trans. Signal Processing*, vol. 44, pp. 281–295, Feb. 1996.

[3] J.-K. Hwang and Y.-C. Chen, "A combined detection-estimation algorithm for the harmonic-retrieval problem," *Signal Process.*, vol. 30, pp. 177–197, 1993.

[4] H.-T. Li and P. M. Djurić, "A novel approach to detection of closely spaced sinusoids," *Signal Process.*, vol. 51, pp. 93–104, 1996.

[5] Y. Bresler and A. Macovski, "Exact maximum likelihood parameter estimation of superimposed exponential signals in noise," *IEEE Trans. Acoust. Speech, Signal Processing*, vol. ASSP-34, pp. 1361–1375, Apr. 1986.

[6] P. Stoica, R. L. Moses, B. Friedlander, and T. Söderström, "Maximum likelihood estimation of the parameters of multiple sinusoids from noisy measurements," *IEEE Trans. Acoust. Speech, Signal Processing*, vol. 37, pp. 378–392, Mar. 1989.

[7] R. Kumaresan, L. L. Scharf, and A. K. Shaw, "An algorithm for pole-zero modeling and analysis," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-34, pp. 637–640, June 1986.

[8] H. Akaike, "Information theory and an extension of the maximum likelihood principle," in *Proc. 2nd Int. Symp. Inform. Theory*, B. N. Petrov and F. Caski, Eds., 1972.

[9] J. Rissanen, "Modeling by the shortest data description," *Automatica*, vol. 14, pp. 465–471, 1978.

[10] P. M. Djuric, "Model selection based on asymptotic Bayes theory," in *Proc. IEEE Seventh SP Workshop Stat. Signal Array Process.*, Quebec City, Quebec, 1994, pp. 7–10.

[11] P. M. Djuric, "A model selection rule for sinusoids white Gaussian noise," *IEEE Trans. Signal Processing*, vol. 44, pp. 1744–1751, July 1996.

[12] G. Schwarz, "Estimating the dimension of a model," *Ann. Stat.*, vol. 6, no. 2, pp. 461–464, 1978.

[13] J.-J. Fuchs, "Estimating the number of sinunoids in additive white noise," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 36, pp. 1846–1854, Dec. 1988.

[14] J. J. Rissanen, "Fisher information and stochastic complexity," *IEEE Trans. Inform. Theory*, vol. 42, pp. 40–47, Jan. 1996.

[15] P. M. Djurić, "Asymptotic MAP criteria for model selection," *IEEE Trans. Signal Processing*, vol. 46, pp. 2726–2735, Oct. 1998.

[16] A. Wald, "Tests of statistical hypothesis concerning several parameters when the number of observations is large," *Trans. Amer. Math. Soc.*, vol. 54, pp. 426–482, Nov. 1943.

[17] P. Stoica, T. Söderström, and V. Šimonytė, "On estimating the noise power in array processing," *Signal Process.*, vol. 26, pp. 205–220, 1992.

[18] G. Xu and T. Kailath, "Fast subspace decomposition," *IEEE Trans. Signal Processing*, vol. 42, pp. 539–551, Mar. 1994.

[19] W. Steedly and R. Moses, "The Cramér-Rao bound for pole and amplitude estimates of damped exponential signals in noise," *IEEE Trans. Signal Processing*, vol. 41, pp. 1305–1318, Mar. 1993.

[20] C. R. Rao, "Information and the accuracy attainable in the estimation of statistical parameters," *Bull. Calcutta Math. Soc.*, vol. 37, pp. 81–91, 1945.

[21] A. Sabharwal and L. Potter, "Wald statistic for selection of nested nonlinear models," submitted for publication.

[22] H. Lee and F. Li, "Quantification of the difference between detection and resolution thresholds for multiple closely spaced emitters," *IEEE Trans. Signal Processing*, vol. 41, pp. 2274–2277, June 1993.

[23] P. Stoica, V. Šimonytė, and T. Söderström, "On the resolution performance of spectral analysis," *Signal Process.*, vol. 44, pp. 153–161, 1995.

[24] E. Dilaveroğlu, "Nonmatrix Cramér-Rao bound expressions for high-resolution frequency estimators," *IEEE Trans. Signal Processing*, vol. 46, pp. 463–474, Feb. 1998.

[25] H. B. Lee, "The Crameŕ-Rao bound on frequency estimates of signals colsely spaced in frequency," *IEEE Trans. Signal Processing*, vol. 40, pp. 1508–1517, June 1992.

[26] D. Starer and A. Nehorai, "Conditional and unconditional maximum likelihood estimation of the parameters of multiple exponential signals in noise," Center Syst. Sci., Yale Univ., New Haven, CT, Tech. Rep. 8914, May 1989.

[27] ——, "Newton algorithms for conditional and unconditional maximum likelihood estimation of the parameters of exponential signals in noise," *IEEE Trans. Signal Processing*, vol. 40, pp. 1528–1534, June 1992.

[28] I. Ziskind and M. Wax, "Maximum likelihood localization of multiple sources by alternating projection," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 36, pp. 1553–1560, Oct. 1988.

[29] C. J. Ying, "Stochastic exponential modeling and application to radar signal processing," Ph.D. dissertation, The Ohio State Univ., Columbus, 1996.

[30] Y. Hua and T. K. Sarkar, "Matrix pencil method for estimating parameters for exponentially damped/undamped sinusoids in noise," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 38, pp. 814–824, 1990.

[31] H. Akaike, "A new look at the statistical model identification," *IEEE Trans. Automat. Contr.*, vol. AC-19, pp. 716–723, Dec. 1974.

[32] P. M. Djuric, "Simultaneous detection and frequency estimation of sinusoids signals," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Minneapolis, MN, Apr. 27–30, 1993, pp. IV53–IV56.

[33] D. J. Andersh, S. W. Lee, and H. Ling, "XPATCH: A high frequency electromagnetic scattering prediction code and environment for complex three-dimensional objects," *IEEE Antennas Propagat. Mag.*, vol. 36, pp. 65–69, Feb. 1994.

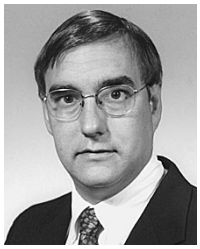
[34] V. U. Reddy and L. S. Biradar, "SVD-based information theoretic criteria for detection of the number of undamped/damped sinusoids and their performance analysis," *IEEE Trans. Signal Processing*, vol. 41, pp. 2872–2881, Sept. 1993.

Ching-Hui J. Ying received the Diploma in electrical engineering from National Taipei Institute of Technology, Taipei, Taiwan, R.O.C. in 1987 and the M.S. degree in 1994 and the Ph.D. degree in 1996, both in electrical engineering, from the Ohio State University, Columbus.

From 1987 to 1989, he was a member of the Military Police in Taiwan. From 1989 to 1990, he was a Research Engineer with the China Town Company. He is now a Buddhist Monk in Taiwan.

Ashutosh Sabharwal received the B.Tech. degree in electrical enegineering from Indian Institute of Technology, New Delhi, in 1993 and the M.S. and Ph.D. degrees, boith in electrical engineering in from the Ohio State University, Columbus, in 1995 and 1999, respectively.

He is currently a Postdocoral Research Associate at Rice University, Houston, TX. His current research interests include statistical signal processing and communications.



Randolph L. Moses (S'78–M'85–SM'90) received the B.S., M.S., and Ph.D. degrees in electrical engineering from Virginia Polytechnic Institute and State University, Blacksburg, in 1979, 1980, and 1984, respectively.

During the summer of 1983, he was a SCEEE Summer Faculty Research Fellow at Rome Air Development Center, Rome, NY. From 1984 to 1985, he was with the Eindhoven University of Technology, Eindhoven, The Netherlands, as a NATO Postdoctoral Fellow. Since 1985, he has been with the Department of Electrical Engineering, The Ohio State University, Columbus, and is currently a Professor there. From 1994 to 1995, he was on sabbatical leave as a Visiting Researcher at the System and Control Group, Uppsala University, Uppsala, Sweden. His research interests are in digital signal processing, and include parametric time series analysis, radar signal processing, sensor array processing, and system identification.

Dr. Moses served on the Technical Committee on Statistical Signal and Array Processing of the IEEE Signal Processing Society from 1991 to 1994. He is a member of Eta Kappa Nu, Tau Beta Pi, Phi Kappa Phi, and Sigma Xi.