Exact Maximum Likelihood Parameter Estimation of Superimposed Exponential Signals in Noise

YORAM BRESLER, MEMBER, IEEE, AND ALBERT MACOVSKI, FELLOW, IEEE

Abstract—A unified framework for the exact maximum likelihood estimation of the parameters of superimposed exponential signals in noise, encompassing both the time series and the array problems, is presented. An exact expression for the ML criterion is derived in terms of the linear prediction polynomial of the signal, and an iterative algorithm for the maximization of this criterion is presented. The algorithm is equally applicable in the case of signal coherence in the array problem. Simulation shows the estimator to be capable of providing more accurate frequency estimates than currently existing techniques. The algorithm is similar to those independently derived by Kumaresan et al. In addition to its practical value, the present formulation is used to interpret previous methods such as Prony’s, Pisarenko’s, and modifications thereof.

I. INTRODUCTION

THE estimation of the parameters of multiple superimposed exponential signals in additive Gaussian noise is of interest in time series analysis and system identification and in antenna array processing. Modern “high resolution” methods (e.g., [1]–[9]) that address variants of this problem provide excellent performance often approaching the Cramer–Rao lower bound at high SNR and with long data records. However, the performance of all methods is severely degraded when the SNR and/or data record length fall below some threshold, which may be significantly higher than that of the optimal maximum likelihood (ML) estimate [4]. This behavior may be attributed to the fact that the above methods are essentially heuristic (“least squares”) modifications of algorithms that yield exact results when there is no noise, or when the amount of available data is infinite (known covariance case). A different class of algorithms based on a model identification approach, e.g., [10]–[13], although using explicitly the correct special ARMA model [14] for exponentials in noise, provides at best asymptotic (for infinite data) approximations to the ML. Furthermore, [10]–[13] address only the time series problem, and not the multieperiment (or array) problem as defined in the next section.

The exact ML estimate is therefore of interest in applications where such suboptimality as displayed by [1]–[13] cannot be tolerated. Unfortunately, direct maximization of the likelihood function, which is highly nonlinear in the unknown signal parameters, requires a computationally expensive multidimensional search. We take here a different route (see also [18]), deriving an exact expression for the ML criterion in terms of the prediction polynomial of the noiseless signal. An iterative algorithm for the minimization of the ML criterion is then presented, which requires only the solution of a quadratic minimization problem at each step and converges in a small number of steps, to produce the ML estimate of the prediction polynomial. By the invariance principle of the ML estimate, the roots of the polynomial are the ML estimates of the signal exponential parameters. An important aspect of our approach is that it is equally applicable in cases of signal coherence in the array problem, which have required specific treatment in previous approaches, such as in the work of Evans et al. [19] and Shan et al. [20].

Almost identical algorithms have been derived independently of ours by Kumaresan et al. [21] and by Kumaresan and Shaw [22], respectively, for the time series and the array problems, by adapting an algorithm for filter design of Evans and Fischl [23]. The differences between our algorithm and theirs are pointed out in Section IV. In particular, we describe an extension of the algorithm to the case when, rather than the signals themselves, only their sample covariance is available.

In addition to the practical value of the algorithm described in this paper, the present formulation is used to interpret previous methods such as the methods of Prony, Pisarenko, and modifications thereof (e.g., [4]–[6], and [9]), as (often poor) approximations to the exact ML; some of them can, in fact, be shown to result from the first iteration of our algorithm. Furthermore, our formulation provides a unifying framework in two respects: 1) complex, real, either damped or undamped sinusoids are treated by the same approach, by imposing various constraints on the solution polynomial; 2) the same formulation encompasses the single-experiment (time series) and the multieperiment (or array) problems.

II. MODEL DEFINITION AND PROBLEM STATEMENT

A set of $N$ measured data vectors ("snapshots") $y_j, j = 1, \cdots, N$, is available. The $m$ components of the $j$th vector $y_j$ are given by

$$y_j(t) = \sum_{i=1}^{p} s_j \lambda_i^j + n_j(t), \quad t = 0, \cdots, m - 1 \quad (1)$$
where $\lambda_i$ is the complex parameter of the $i$th signal, assumed distinct from the other signals, i.e., $\lambda_i \neq \lambda_j$, $\forall i \neq j$, and $s_j$ is the complex amplitude of the $i$th signal in the $j$th snapshot, which may vary across $j$. $n_j(t)$ is a complex normal random variable uncorrelated across both $t$ and $j$, with uncorrelated real and imaginary components, each of variance $\sigma^2/2$.

The model is compactly described by the following vector notation. Each vector is length $m$ given by the sum of $p < m$ exponential signal vectors corrupted by additive noise,

$$y_j = \sum_{i=1}^{p} s_j \ a(\lambda_i) + v_j \quad (2)$$

where

$$a(\lambda_i) = [1, \lambda_i, \lambda_i^2, \ldots, \lambda_i^{q-1}]'. \quad (3)$$

Here, $v_j$, $j = 1, \ldots, N$, is a complex, white Gaussian zero mean vector random process with uncorrelated real and imaginary components, and a scaled identity covariance matrix 4

$$E v_j v_j^H = \sigma^2 \cdot I \quad (4)$$

with $\sigma^2$ possibly unknown. We use $(\cdot)^H$ to denote the Hermitian-transpose operation, and $(\cdot)^*$ to denote the complex conjugate. Defining the $m \times p$ Vandermonde matrix

$$A(\lambda) = [a(\lambda_1), a(\lambda_2), \ldots, a(\lambda_p)] \quad (5)$$

where $\lambda = [\lambda_1, \lambda_2, \ldots, \lambda_p]'$ is the signal parameter vector, and the $j$th snapshot signal-amplitude-vector is $s_j = [s_{1j}, s_{2j}, \ldots, s_{pj}]'$, (1) can be rewritten as

$$y_j = A(\lambda) \ s_j + v_j \quad (6)$$

The above formulation encompasses the data model for a variety of problems. The case $N = 1$ corresponds to the single-experiment time series problem with uniform sampling, the components of the vector $y = y_1$ being the $m$ time samples of the measured signal. The case $N > 1$ corresponds to a multiple experiment with a time series, or equivalently, to the data model for a linear uniform narrow-band array, with multiple plane waves (far-field sources) present. In the latter case, the components of the measurement vector (the “snapshot”) represent the output from $m$ individual sensors. Further specializations are obtained by the choice of $\lambda_i$ and $s_j$ as follows.

a) Damped complex sinusoids. $\lambda_i = \exp (j \omega_i - \alpha_i)$, where $0 \leq \omega_i \leq \pi$ and $0 \leq \alpha_i$ are, respectively, the frequency of the sinusoid and its damping coefficient.

b) Pure complex sinusoids. For every $i$, $\alpha_i = 0$ in a) above.

c) Damped real sinusoids. $p = 2q$ (even), $\lambda_i = \exp (j \omega_i - \alpha_i)$, $\lambda_{p-i} = \lambda_i^*$, and $s_{p-i} = s_{q-i}^*$, $i = 1, \ldots, q$.

d) Pure real sinusoids. For every $i$, $\alpha_i = 0$ in c) above.

Note that the array problem is described by case b) and

$N \geq 1$, by making the identification $\omega_i = k_i \Delta$. Here $k_i$ is the wavenumber of $i$th plane wave impinging on the array, and $\Delta$ is the element spacing.

We address the general problem (5), (6) with arbitrary exponential signals. The specialization to individual cases will be achieved by appropriately constraining the solution.

Assuming that the number of signals $p$ is known 5 the estimation problem is

Given the data $\{y_j\}_{j=1}^{N}$,

estimate the signal parameter vector $\lambda$ \hspace{1cm} (A)

estimate the signal amplitudes $\{s_j\}_{j=1}^{N}$. \hspace{1cm} (B)

In other work on the array problem, the additional modeling assumption that the amplitudes $s_j$ are i.i.d. (independent identically distributed) samples of a zero mean Gaussian vector with a covariance matrix $S = E[s_j s_j^H]$ is often added. When only the directions of arrival and a “source signature” are of interest, (B) is replaced by

estimate the signals amplitude covariance matrix $S$. \hspace{1cm} (B')

It appears that in most applications there is little support for this additional assumption, e.g., there is no reason to assume that the amplitudes of signals received in an array from different sources should be jointly normally distributed. The same remark applies to multiple experiments on a time series. Therefore, in this paper, we relax any distribution assumption on the $s_j$, and proceeding in the only logical way, consider them to be unknown constant vectors 6 (possibly different for different $j$), to be estimated similarly to $\lambda$.

III. THE MAXIMUM LIKELIHOOD CRITERION

Under the white Gaussian noise (WGN) assumption, the ML estimate of the signal parameters $\lambda$ and amplitudes $\{s_j\}_{j=1}^{N}$ is obtained by solving the nonlinear least squares problem

$$\min_{\lambda \in \Lambda} \sum_{j=1}^{N} \sum_{s \in D} \| y_j - A(\lambda) s_j \|^2 \quad (7)$$

where $\Lambda$ and $D$ represent, respectively, the sets of values $\lambda$ and $\{s_j\}_{j=1}^{N}$ can take under the constraints of the specific model assumed—e.g., one of a) + d) above, and where $\| \cdot \|$ represents the Euclidean norm.

Significant computational savings follow from the observation [25] that for any given $\lambda$ the problem of finding the optimal $\{s_j\}_{j=1}^{N}$ corresponding to that $\lambda$, is a linear least squares problem whose solution is given by

$$s_j = A^*(\lambda) y_j \quad (8)$$

$^5$The case of unknown number of signals can be dealt with, e.g., by information theoretic criteria as in [24].

This choice eliminates the dependence of the estimation procedure on signal-to-signal correlation and, in particular, the difficulty that all high resolution methods experience with coherent sources does not arise.
where \( A^\dagger(\lambda) = [A^H(\lambda) A(\lambda)]^{-1} A^H(\lambda) \) is the pseudoinverse of \( A(\lambda) \). By substituting (8) into (7), the \( s_i \) are eliminated, and the problem (7) is reduced (see [25], [7], [21], [22] for details) to one of the equivalent formulations
\[
\min_{\lambda \in \Lambda} J_1(\lambda), \quad J_1(\lambda) = \text{tr} \left[ P_{\lambda}(\lambda) \hat{R}_y \right]
\]
where \( P_{\lambda}(\lambda) \) and \( P_{\lambda}^\dagger(\lambda) \) are, respectively, the projection matrices onto the column space of \( A(\lambda) \) and onto its orthogonal complement,
\[
P_{\lambda}(\lambda) = A(\lambda) A^\dagger(\lambda)
\]
\[
P_{\lambda}^\dagger(\lambda) = I - P_{\lambda}(\lambda),
\]

\[\text{tr} \] is the trace operator, and
\[
\hat{R}_y \triangleq \frac{1}{N} \sum_{j=1}^{N} y_j y_j^H
\]
is the sample autocorrelation matrix of \( y_j \). Once the ML estimate \( \hat{\lambda}_{ML} \) is determined by solving either (9a) or (9b), \( \{s_j\}_{j=1}^{N} \) are found by the linear relationship (8).

We now make the following observations. 1) The measurement sample correlation \( \hat{R}_y \) as defined by (12) (see our discussion below of other sample correlations) is sufficient for the ML estimation of \( \lambda \). 2) \( \hat{R}_y \) is not sufficient for the ML estimation of \( \{s_j\}_{j=1}^{N} \) : all the measurements \( \{y_j\}_{j=1}^{N} \) are necessary to determine \( \{s_j\}_{j=1}^{N} \). However, if only an estimate of the sample correlation
\[
\Sigma \triangleq \frac{1}{N} \sum_{j=1}^{N} s_j s_j^H
\]
of the \( s_j \) is required, \( \hat{R}_y \) is again sufficient, since by the invariance principle of the ML estimator it follows using (8) that
\[
(\Sigma)_{ML} = \frac{1}{N} \sum_{j=1}^{N} s_j s_j^H = \frac{1}{N} \sum_{j=1}^{N} A^\dagger(\lambda) y_j y_j^H A^H(\lambda)
\]
\[
= A^\dagger(\lambda) \hat{R}_y A^H(\lambda).
\]

Thus, if we wish to obtain only \( \hat{\lambda} \) and \( \Sigma \), \( \hat{R}_y \), which has a lower dimensionality than \( \{y_j\}_{j=1}^{N} \) whenever \( N > m \) [number of snapshots greater than \( \text{dim} (y_j) \)], may be used rather than the complete data set, effectively performing “data compression” and leading to a reduction in computation.

The criteria (9) will be, in general, multimodal, rendering the multidimensional search for a global extremum computationally expensive. Therefore, in [7] and [26], the eigenstructure decomposition of the sample correlation matrix is used to derive interesting 1-D search approximations. In this paper we take a different approach, avoiding altogether the eigendecomposition of the signal correlation matrix.

IV. THE ML CRITERION FOR THE MODEL PARAMETERS

The ML criteria (7) and (9) are expressed in terms of the signal parameters. Past work on exponential signal estimation (in the single experiment case), however, has often been in terms of the identification of a model for the signals. The motivation is that the components \( y_j(t) \) of \( y_j \) in (1) obey the “special ARMA model” [14]
\[
b_0 y_j(t) + b_1 y_j(t - 1) + \cdots + b_p y_j(t - p) = b_0 \eta_j(t) + b_1 \eta_j(t - 1) + \cdots + b_p \eta_j(t - p)
\]
where the polynomial
\[
b(z) = b_0 z^p + b_1 z^{p-1} + \cdots + b_p
\]
is the linear prediction (LP) polynomial for the noiseless signal, and has roots at \( z_i = \lambda_i \), \( i = 1, \cdots, p \). If the coefficient vector \( b = [b_0, b_1, \cdots, b_p] \) is estimated by some identification method \( b(z) \) is often chosen as monic, i.e., \( b_0 = 1 \). If the root of \( b(z) \) will provide estimates of \( \{\lambda_i\}_{i=1}^{p} \) (i.e., \( \lambda \)). As indicated in the Introduction, all identification methods used so far for this problem, have been either heuristic modifications of schemes for the noiseless case or, at best, asymptotic approximations to the ML. An exact expression for the ML criteria in terms of the model parameters, e.g., \( b_1 \cdots b_p \) appears to be difficult to obtain from (16) and has been derived only recently by the authors [18]. The least squares criterion derived independently in [21] and [22], although not couched in statistical terms, is essentially equivalent. We derive now such an expression, using (9), and its geometrical interpretation.

**Claim:** Let \( B \) be the \( m \times (m - p) \) Toeplitz matrix
\[
\begin{bmatrix}
b_0 & 0 & \cdots & 0 \\
0 & b_0 & \cdots & 0 \\
& & \ddots & \ddots \\
& & & b_0
\end{bmatrix}
\]

Then \( P_B(\lambda) \) can be written as
\[
P_B(\lambda) = B(B^H B)^{-1} B^H.
\]

**Proof:** Let \( a(\lambda_i) \) be a column of \( A(\lambda) \). Then for \( i = 1, \cdots, p, \)
\[
B^H a(\lambda_i) = (a(\lambda_i) \begin{bmatrix}1, \lambda_i, & \cdots, \lambda_i^{n-p-1} \end{bmatrix})^T = \mathbf{O},
\]
where the last equality follows from the fact that \( b(\lambda_i) = 0 \). It follows that the columns of \( B \) are orthogonal to those of \( A \). Now, since for any \( b \neq \mathbf{O}, B \) has full rank equal to \( m - p \), its columns span the orthogonal complement to the range space of the \( p \) columns of \( A \), and the projection \( P_B \) onto the subspace is equal to \( P_B(\lambda) \).
Relation (9a) can now be rewritten as

\[
\min_{b \in \Theta} J(b),
\]

(20a)

\[
J(b) = \text{tr} [P_y \hat{R}_y] = \text{tr} \left[ B (B^H B)^{-1} B^H \hat{R}_y \right],
\]

(20b)

where we leave the constraint set \( \Theta \) unspecified for the time being. Once \( \hat{R}_y \) is obtained by solving (20), \( \lambda \) is obtained by rooting \( b(z) \).

The criterion in (20b) can be further manipulated into a form that is both useful for the interpretation of previous work and for actual computation. First note the following equality (which amounts to a statement of the commutativity of the convolution operation):

\[
B^H y_j = y_j b,
\]

(21)

where the "data matrix" \( Y_j \) is defined as

\[
Y_j = \begin{bmatrix}
y_j(p + 1) & y_j(p) & \cdots & y_j(1) \\
y_j(p + 2) & y_j(p + 1) & \cdots & y_j(2) \\
\vdots & \vdots & \ddots & \vdots \\
y_j(m) & y_j(m - 1) & \cdots & y_j(m - p)
\end{bmatrix}.
\]

(22)

Substituting (12) back into (20b), and using this result, we obtain

\[
N \cdot J(b) = \text{tr} \left[ \sum_{j=1}^{N} B (B^H B)^{-1} B^H y_j^H \right] = b^H \left[ \sum_{j=1}^{N} Y_j^H (B^H B)^{-1} Y_j \right] b.
\]

(23)

Given Sample Correlation: A similar result can be obtained in terms of \( \hat{R}_y \), by considering any of its square root factors, e.g., its Cholesky factors \( L \) and \( L^H \), where \( L \) is lower triangular. Let \( I_j \) be the \( j \)th column of \( L \), then

\[
\hat{R}_y = L L^H = \sum_{j=1}^{m} I_j I_j^H.
\]

(24)

Substituting (24) into (20), and going through the operations as in (23), yields

\[
J(b) = b^H \left[ \sum_{j=1}^{m} L_j^H (B^H B)^{-1} L_j \right] b,
\]

(25)

where \( L_j \) is constructed from \( I_j \) in the same manner as \( Y_j \) is from \( y_j \).

We proceed now to complete the specification of the ML criterion (23) by defining the constraint set \( \Theta \).

Constraint Sets \( \Theta \)

The constraint \( b \in \Theta \), which has been left unspecified so far, is equivalent to \( \lambda \in \Lambda \) in (9a) and is constructed as the intersection of two constraints

\[
\Theta = \Theta_0 \cap \Theta_r.
\]

(26)

The "nontriviality constraint"

\[
\Theta_0 \triangleq \{ b : b \neq 0 \}
\]

(27)

makes the correspondence between \( b \) and \( \lambda \) unique, at the same time ensuring the invertibility of \( (B^H B) \) for all \( b \in \Theta \), owing to the Toeplitz structure of \( B \). Possible examples are \( b_0 = 1 \) [monic \( b(z) \)], \( \| b \| = 1 \), etc. The constraint set \( \Theta_r \), \( r = 1, \ldots, 4 \) is chosen to complete the correspondence to the respective constraint on \( \lambda \) of Section II, whose use to constrain the minimization in (23) or (25) is part of the ML formulation, and is essential for optimality, as follows:

- a) \( \Theta_1 \triangleq \{ b : b(z) \text{ has roots inside or on the unit circle} \} \),
- b) \( \Theta_2 \triangleq \Theta_1 \cap \{ b : b_i = b_{p-i}, i = 0, \ldots, p \} \),
- c) \( \Theta_3 \triangleq \Theta_1 \cap \{ b : \text{Im}(b) = 0 \} \),
- d) \( \Theta_4 \triangleq \Theta_2 \cap \Theta_3 \).

(28)

Note that (28b) no longer holds in general if \( b(z) \) is forced to be monic. Accordingly, we avoid the latter constraint in the case of pure complex sinusoids. The symmetry constraints in (28) may be formulated explicitly as linear or quadratic constraints on \( b \), or introduced in the construction of \( Y_j \). The latter approach is preferred in this paper (see Section VI below), as it reduces the dimensionality of the problem.

An Interpretation of the ML Criterion

For simplicity, consider the case \( N = 1 \). Relation (23) is then reduced to

\[
J(b) = b^H y_H (B^H B)^{-1} Y b.
\]

(29)

Since \( b(z) \) is the prediction-error polynomial, \( e = Y b \) is the prediction-error vector (up to a multiplicative constant introduced by \( b(z) \) being possibly not monic) whose components are the right-hand side of (15) at different \( t \). As noted in Section II, \( e \) is an MA process with coefficients \( b_i \). Its covariance is therefore \( (B^H B) \), so that \( w = (B^H B)^{-1/2} Y b \) is a vector of white noise samples, distributed as \( w \sim N(0, a^2 \cdot I) \). The negative-log-likelihood function for \( w \) is therefore given (up to multiplicative and additive constants) by (29).

Could this reasoning be used to directly obtain (29) as the exact ML criterion? The answer does not seem to be obvious. Recall that (29) is not the negative-log-likelihood for \( y \), but rather its maximum over all possible signal amplitude vectors \( s \). Also, the transformation from \( y \) to \( w \) is not 1:1, so that the conditional density of \( y \) cannot be written only in terms of the density of \( w \).

V. PREVIOUS METHODS AND THE EXACT ML CRITERION

In this section we use the formulations (20a) and (23) of the ML criterion and accompanying constraints (29) to obtain some insight into approaches proposed in previous work on the resolution of sinusoids in noise.

Constraint Sets and "High Resolution"

A formulation of the array estimation problem as an optimization problem with linear or quadratic constraints
has been previously suggested by Bronez and Cadzow [8]. The criterion to be minimized is derived there by heuristic arguments extending the noiseless case and, unlike the optimal ML (23), it is quadratic in \( b \). The constraints considered in [8] are only those corresponding to our "non-triviality constraints" guaranteeing \( b \neq 0 \), and the authors study the choice of constraint in this class to optimize "resolution." Clearly, the ML criterion (23) eliminates the need for such study, since the resulting estimate is the same for any of these (or other) constraint choices corresponding to (27), true resolution being determined by other factors as we elaborate below.

The use of constraints such as in (28) (in particular, (28d) for real sinusoids) has been suggested previously (cf. [9]–[11] and [27]) to ensure parsimony and reduce estimation error. However, it appears that in our formulation, which is posed as a constrained optimization problem, the full generality is easier to appreciate. For example, the cases of pure complex sinusoids and of damped complex sinusoids are treated separately by Tufts and Kumaresan by different methods—respectively, in [4] and [5]. We note that relaxing the constraint on \( b(z) \) to be monic, and applying (28b) to the normal equations obtained in [4] by the forward–backward "least squares" method of Ulrych and Clayton [14], results in the elimination of the "backward" equations which become linearly dependent on the "forward" ones. Thus, both approaches can be made equivalent. However, the explicit use of (28b) provides parsimony resulting in reduced problem size.

The present formulation also has the advantage that it decouples the issues of "high resolution," "spectral smoothing," and estimation accuracy. Spectral smoothing (see, e.g., Kay and Marple [27]) has been often considered the culprit for "low resolution—the broadening of spectral lines in the estimate to the point where they could not be resolved (by eye). Spectral smoothing, however, can be avoided in any method (even with linear prediction) by the use of a correct constraint on \( b(z) \). For example, employing (28b), the zeros of \( b(z) \) are guaranteed to lie on the unit circle, producing spectral lines in the "spectrum" \( 1/|b(e^{j\omega})|^2 \). Thus, the resolution has little to do with the absence of spectral smoothing; it is determined by the detection performance of the algorithm in not underestimating the number of signals present even when their parameters are close. (The detection problem is not dealt with here, as \( p \) is assumed known.) If the number of signals is correctly estimated, e.g., by a hypothesis testing procedure (or over estimated as is common practice in LP methods), the remaining issue is estimation accuracy; having the spectral lines appear at the correct frequency. Both detection and estimation performance depend, of course, on the \( S/N \) and the specific method employed. The need for a distinction between apparent resolution (as determined by, e.g., lack of spectral smoothing) and detection performance, has also been pointed out by Kay and Demeure [28].

**The Pisarenko Method with Sample Covariance**

Rife and Boorstyn [29] have shown in the time series case that a DFT-based spectral estimator is a realization of the exact MLE for a single pure sinusoidal signal in white noise \( (p = 1) \). The corresponding result for the array case obtained by Wax [7] is that classical beamforming is equivalent to the exact MLE for a single plane wave in spatially white noise \( (p = 1) \). We show now a similar result for the Pisarenko method when \( p = m = 1 \).

When \( p = m = 1 \), we have \( B = B^H \), and (20) becomes

\[
\min_{b \in \mathbb{C}^m} b^H \hat{R}_p b \quad (30)
\]

Without constraints, when \( \Theta = C^{p+1} \), (30) evaluates to \( \| \hat{R}_p \|_F \)—the \( L_2 \) norm of \( \hat{R}_p \), and the minimizing \( b \) is the eigenvector of \( \hat{R}_p \), corresponding to its minimum eigenvalue. This is exactly the estimate provided by the Pisarenko method for harmonic retrieval [1] when applied to sample covariance of a signal.

We conclude, therefore, that the Pisarenko method coincides with the ML if and only if \( p = m = 1 \) and no a priori information on the pole location of \( b(z) \) is available. (Otherwise, constraints should be applied, modifying the method.) Clearly, the last requirement is not met in the application for which the Pisarenko method has been designed: harmonic retrieval where the roots are known a priori to be on the unit circle. The Pisarenko method is therefore never optimal.

**Constrained Pisarenko Method**

If and only if \( p = m = 1 \), the Pisarenko method for a sample covariance can be modified to coincide with the ML by employing any of the constraints in (28) as appropriate for the specific case at hand. For example, with pure complex sinusoids, the constraint (28b) may be stated as a quadratic constraint on \( b \) and incorporated into (30) using a Lagrange multiplier, resulting in (30) being modified into a generalized eigenvalue problem. Alternatively, \( \hat{R}_p \) may be modified to account for the symmetry of \( b \), as was done by Marple [9] for the Prony method with real sinusoids, or in the next section for complex sinusoids.

The application of symmetry constraints in the context of the Pisarenko method, as suggested here, appears to be new. Moreover, we have established, under appropriate conditions, the equivalence of the constrained Pisarenko method to the optimal ML estimator with given finite data.

**Prony’s Method and “Spatial Smoothing”**

Consider first the time series case \( N = 1 \), where Prony’s method has been traditionally applied. The cost to be minimized is given in (29). If we replace \( (B^H B) \) by the identity matrix, we obtain

\[
\min_{b \in \mathbb{C}^m} b^H y^H y b \quad (31)
\]
If, in addition, we choose as the only constraint the non-triviality constraint "b(z) is monic," i.e., we set \( \Theta = \Theta_{b} = \{ b : b_{0} = 1 \} \), we obtain exactly the criterion for linear prediction by the covariance method, or Prony’s method for the fitting of exponentials as extended by Hildebrand (cf. [6]) with prediction order set equal to the number of signals. Constraints from (28) may be incorporated similarly to the discussion above.

Is there a situation where Prony’s method coincides with the ML? The answer is: not in any useful case. For clearly, with \( b(z) \) monic and when \( m \geq 2p \) [as required for a unique solution of (31)], the relation \( B^{H}B = I \) holds if and only if \( b = [1, 0, \cdots, 0]' \), which does not correspond to exponential signals.

Similar conclusions apply to the multiexperiment case where (31) is replaced by its obvious generalization to

\[
\min_{b} b^{H}K_{r}b, \quad K_{r} \triangleq \sum_{j=1}^{N} Y_{j}^{H}Y_{j}. \tag{32}
\]

The use of the matrix \( K_{r} \) to combat the difficulties of high resolution methods in the presence of signal coherence has been motivated by a ‘‘spatial smoothing’’ argument by Evans et al. [19] and by Shan et al. [20]. They consider (in the array problem) ‘‘subarrays’’ formed by selecting all \( m = d + 1 \) possible groups of \( d \) consecutive sensors from the array, and calculate the individual sample correlations of the corresponding data vectors. Then, they average these sample correlations to form the ‘‘spatially smoothed’’ sample correlation which, for \( d = p + 1 \), is exactly \( K_{r} \).

Comparing (29) and (31) [respectively, (23) and (32)], it becomes obvious that the ‘‘smoothed sample correlation’’ \( Y_{j}^{H}Y \) (respectively, \( 1/N \sum_{j=1}^{N} Y_{j}^{H}Y_{j} \)) is insufficient for the determination of the MLE of the signal parameters so that any method using it as the sole information is bound to be inferior to the exact ML. Thus, the need to use a prediction order much in excess of the true number of signals present in order to improve the performance of the Prony method (cf. Kumaresan et al. [6]) can be explained not only as a way to obtain a better fit to the ‘‘special ARMA’’ by an AR model, but also by the need to preserve information in the ‘‘sample correlation’’ \( Y_{j}^{H}Y \): the higher the prediction order, the closer is \( Y_{j}^{H}Y \) to \( \hat{R}_{j} \).

Before leaving Prony’s method, we note that it may be applied not only to actual signals but also the sample correlation matrix \( \hat{R}_{j} \), since the ‘‘spatial smoothing’’ operation and the summation across \( j \) commute [19]. This is particularly useful when \( m << N \) (‘‘data compression’’ and reduced computation) or when only \( \hat{R}_{j} \) is available. A form employing the square root factors of \( \hat{R}_{j} \) may be obtained by setting \( B^{H}B = I \) in (25), which yields

\[
\min_{b} b^{H} \left( \sum_{j=1}^{m} L_{j}^{H}L_{j} \right) b. \tag{33}
\]

VI. AN ML ESTIMATION ALGORITHM

The ML estimate of \( b \) (and of \( \lambda \)) is obtained by solving (20a), (23), or (20a), (25) subject to one of (28a)–(28d). This is a nonlinear minimization problem with linear or quadratic constraints, which is not obviously easier than problem (9)–the one posed directly in terms of \( \lambda \). However, we now present an iterative algorithm for the required constrained minimization which requires only the solution of a quadratic minimization problem in each step and converges in a small number of steps (typically less than 10).

**Basic Algorithm**

The proposed algorithm, termed IQML (iterative quadratic ML), consists of the following steps, iterated until a convergence criterion is met. (\( \varepsilon \) is chosen according to the desired precision.)

a) **Initialization:** \( k = 0 \) and \( b_{(0)} = b_{0} \).

b) **Compute:**

\[
C_{Y}^{(k)} = \sum_{j=1}^{N} Y_{j}^{H} \left( B_{(k)}^{H}B_{(k)} \right)^{-1}Y_{j} \tag{34a}
\]

or

\[
C_{Y}^{(k)} = \sum_{j=1}^{m} L_{j}^{H} \left( B_{(k)}^{H}B_{(k)} \right)^{-1}L_{j} \tag{34b}
\]

where \( L_{j} \) are as defined in (25), and \( B_{(k)} \) is constructed from \( b_{(k)} \) in the same way as \( B \) in (17) is from \( b \).

c) **Solve the Quadratic Minimization Problem:**

\[
\min_{b} b^{H}_{(k+1)}C_{Y}^{(k)}b_{(k+1)} \tag{35}
\]

where \( \Theta \) is chosen in accordance to the appropriate case from (26)–(28).

d) \( k = k + 1 \).

e) **Check Convergence:** \( \| b_{(k)} - b_{(k+1)} \| < \varepsilon \)? (Yes—done, go to f). No—go to b).

f) **Find the Roots of** \( b_{(k)}(z) \): These are the signal parameters \( \lambda_{i} \), \( i = 1, \cdots, p \).

**Implementation of Constraints**

**Nontriviality Constraint:** Since the specific choice of \( \Theta_{b} \) does not affect the final result, we have chosen for convenience \( b_{0} = 1 \) [or Re \( b_{0} \) = 1 for complex sinusoids].

**Stability of** \( b(z) \): In general, the solution of the normal equations resulting from (35) may yield an unstable polynomial \( b_{(k)}(z) \) so that satisfaction of (28a) is not automatically guaranteed. Nonetheless, it can be shown that when the SNR is not too low, \( b_{(k)}(z) \) will be stable with high probability, this probability increasing with increasing number of data points or SNR. This is supported by simulation experiments, where \( b_{(k)}(z) \) was stable whenever the SNR or data record length were sufficient to allow above-threshold estimation performance.

**Symmetry of** \( b(z) \): The constraint (28b) or (28d) may be explicitly formulated as a quadratic constraint [10], or implemented implicitly when forming \( Y_{j} \). For pure real sinusoids, the latter implementation is straightforward as shown, e.g., in [9]. We consider here in some detail the...
pure complex sinusoid case which is used in the simulation. The implicit implementation of a symmetry constraint preferred here converts the problem into an unconstrained one and reduces its size by half.

Let \( p = 2q + 1 \) be odd. (The even case is similar.) Partition \( b \) and \( Y \) such that

\[
 b = [b_0 \cdots b_q b_{q+1} \cdots b_{2q+1}]' = [b_1' \ b_2']' \tag{36}
\]

and

\[
 Yb = [Y_1 \ Y_2] \begin{bmatrix} 
 b_1 \\
 b_2
\end{bmatrix} = Y_1 b_1 + Y_2 b_2. \tag{37}
\]

Defining the reverse permutation matrix \( \tilde{I} \),

\[
 \tilde{I} \triangleq \begin{bmatrix} 
 0 & 1 \\
 1 & 0 \\
 \vdots & \vdots \\
 1 & 1
\end{bmatrix}, \tag{38}
\]

the symmetry relation in (28b) may be represented by \( b = \tilde{b}^* \) or \( b_2 = \tilde{b}^* \). Defining now

\[
 c \triangleq [\text{Re} \ (b_1'), \ \text{Im} \ (b_1')] \tag{39}
\]

and

\[
 \tilde{Y} = [Y_1 + Y_2 \tilde{I}] [(Y_1 - Y_2 \tilde{I})]. \tag{40}
\]

The criterion becomes

\[
 \min_{c \in \mathbb{R}^{q+1}} \min_{c_0 = 1} c' \Bigg[ \sum_{j=1}^{N} \tilde{Y}^H (B^H B)^{-1} \tilde{Y} \Bigg] \tag{41}
\]

This may be verified to be equivalent to the problem

\[
 \min_{c_0 = 1} c' C_Y c \tag{42a}
\]

with

\[
 C_Y = \sum_{j=1}^{N} \text{Re} \left\{ \tilde{Y}^H (B^H B)^{-1} \tilde{Y} \right\}, \tag{42b}
\]

whose solution is given by the solution to the normal equations

\[
 C_Y c = [1, 0, \ldots, 0]' \tag{43}
\]

Once \( c \) is found, \( b \) is constructed from it in the obvious way. The constraint (28b) is therefore implemented by forming \( C_Y \) in step b) of the algorithm as in (42b) (with the obvious equivalent for \( L_j \)), and by solving the normal equations (43) in step c).

Remarks

1) The matrices \((B^H B)\) and \(C_Y\) have a great deal of structure, which can be taken advantage of, for efficient computation. For example, the inverse of \((B^H B)\), which is a banded Toeplitz matrix, can be computed efficiently employing ideas from [30] or by a decomposition involving a circulant matrix as in [21]. Alternatively, in practice, \((B^H B)^{-1}\) need not be computed explicitly, but rather, a whitening filter can be constructed to implement the operation \( (B^H B)^{-1/2} Y \).

2) The approach is equally applicable in cases of signal coherence in the array problem, which have required specific treatment in previous approaches, such as in the work of Evans et al. [19] and Shan et al. [20].

A criterion similar to our relation (23) (with \( N = 1 \) and without any constraints) occurs in a filter design problem considered earlier by Evans and Fischl [23]. Although derived independently, it turns out that our algorithm coincides (except for the constraints and \( N \geq 1 \)) with the first phase of their algorithm. Also, except for the way the constraints are implemented ([21] and [22] are not specific on this point), our algorithm is identical to the first phase in the algorithms of [21] and [22], which are based on [23]. The second phase of the above-mentioned algorithms involves a gradient-based nonlinear programming algorithm to locate the exact minimum of the criterion with respect to \( b \). It is claimed in [21] and [22] that this may be necessary since, in the first phase, \((B^H B)^{-1}\) is assumed constant at each of the iterative steps, and the fixed point of the iteration may not correspond exactly to the minimum. In our experiments, we have not found this to be a problem (see the simulation results.)

Relation to Other Iterative Algorithms

In addition to [21] and [22], which we have already discussed, Matausek et al. [12] and Kay [13] have suggested similar iterative procedures to estimate pure sinusoids in noise. A fundamental difference from the approach in this paper is that those schemes are derived from transfer function considerations appropriate for a stationary case, which do not apply to the finite data case even when the underlying model is stationary. Therefore, for any finite data record, those methods are suboptimal with departure from optimality increasing with decreasing data record length. Thus, instead of computing \((B^H B)^{-1/2} Y\), Kay applies the constant coefficient “prewhitening” filter \(1/b(z)\) to the \( y = y_i\) vector. Since the poles of this filter approach the unit circle, the effect of unknown initial conditions does not die out.

An additional difference is that [13] does not implement the constraints proposed here on the solution polynomial, whose use may significantly contribute to a decrease in estimation error. Finally, these previous methods address only the single experiment case, the extension to the multiexperiment case not being obvious.

We note that of the three different algorithms suggested in [12], the GLS (which is identical to [13]) is most closely related to our algorithm. The ITIF and the “ML algorithm” of [12] have the advantage that colored noise with unknown covariance can be accommodated. (As indicated above, colored noise with a known covariance can be simply accommodated in our approach.)

Simulation Example

In order to facilitate comparison to previous published work, an example common to [4], [6], [13], and [21] was
chosen for simulation. A single experiment case is considered. Twenty-five data points were generated according to the model

\[ y_t = \exp(j2\pi f_1 t + j\pi/4) + \exp(j2\pi f_2 t) + n_t \quad t = 1, \ldots, 25 \]

with \( f_1 = 0.52, f_2 = 0.5, \) and \( n_t \) a white Gaussian random complex sequence, with uncorrelated real and imaginary parts, each with a variance of \( \sigma^2 \), such that the SNR is 10 log \((1/2\sigma^2)\). The IQML algorithm was initialized with \( \langle B^n_0B_0 \rangle = I \) so that the first iteration of our algorithm results in the Prony estimate.

The MSE results of Monte Carlo simulation (20 independent trials) are shown in Fig. 1, along with the C–R bound for unbiased estimation, and the performance curves for other methods which were extracted from [4] and [13]. Within the finite sample experimental error, the estimates appear unbiased, and the MSE at SNR’s higher than 3 dB coincides with the C–R bound. Below a threshold of about 3 dB, the MSE increases rapidly due to the algorithm occasionally converging to a local rather than global minimum of the ML criterion. At the low SNR end, the proposed algorithm extends the threshold by 4 dB below that of the “modified FBLP method” of Tufts and Kumaresan [4], and by 1–1.5 dB below that of the “iterative filtering algorithm” (“IFA”) of Kay [13]. An improvement over the IFA of Kay is also the achievement of the C–R bound at high SNR’s; the failure of Kay’s IFA to achieve the C–R at high SNR stems from the bias in its estimate, which may probably be attributed to the inappropriateness of stationary modeling at a record length of 25.

The results are similar to those of [21], except for the slightly higher variance above the threshold (which is, however, within the experimental error of our 25 Monte Carlo runs sample), and the location of the threshold which is about 1 dB lower in our results. It therefore appears to us that the addition of the second phase in [21] and [22], with the purpose of setting the derivative of the criterion exactly to zero, does not yield significant improvement over the IQML; the estimation error at convergence is probably dominated by the noise (at least in the range of SNR’s tested), and not by the precision with which the minimum of the criterion has been located.

VII. CONCLUSIONS

An exact expression for the ML criterion for the estimation of the parameters of superimposed exponential signals in noise has been derived in terms of the prediction polynomial of the signal. A unified formulation as a constrained optimization problem is thus obtained for the various cases of exponential signals, encompassing both the time series and the uniform linear array problems. New insight on previous methods may be gained in this framework: a (new) constrained version of the Pisarenko method for the sample covariance was shown to coincide with the MLE when the number of complex sinusoidal signals present is exactly one less than the dimension of the signal vectors (number of sensors in the array application); Prony’s method was shown to never coincide with the MLE. Its extension to the multiexperiment case, and to use the signal’s sample covariance in the case when only the later is available, rather than the complete batch of “snapshots,” was shown to follow naturally as a first iteration of our proposed algorithm.

Finally, an iterative algorithm, IQML, has been proposed for the solution of the constrained optimization problem associated with the MLE requiring only the solution of linear equations. The algorithm is applicable also in the case of “signal coherence,” which is of interest in array processing. Although its convergence properties require further theoretical and experimental investigation, it has been shown by simulation to provide better estimation accuracy than other recently proposed algorithms. Its superior accuracy at short data records suggests that it may be used for the estimation of directions of arrival of multiple sources in a linear array from as little data as a single snapshot.

ACKNOWLEDGMENT

The first author would like to acknowledge the stimulating discussions on eigenstructure methods of array processing with M. Wax, who provided him with a draft of parts from [8], where the ML criterion (9) is derived. Valuable comments by R. Roy and by the reviewers of this paper are gratefully acknowledged.

REFERENCES


