

Amplitude Estimation of Sinusoidal Signals: Survey, New Results, and an Application

Petre Stoica, *Fellow, IEEE*, Hongbin Li, *Member, IEEE*, and Jian Li, *Senior Member, IEEE*

Abstract—This paper considers the problem of amplitude estimation of sinusoidal signals from observations corrupted by colored noise. A relatively large number of amplitude estimators, which encompass least squares (LS) and weighted least squares (WLS) methods, are described. Additionally, filterbank approaches, which are widely used for spectral analysis, are extended to amplitude estimation. More exactly, we consider the recently introduced matched-filterbank (MAFI) approach and show that by appropriately designing the prefilters, the MAFI approach to amplitude estimation includes the WLS approach. The amplitude estimation techniques discussed in this paper do not model the observation noise, and yet, they are all asymptotically statistically efficient. It is, however, their different finite-sample properties that are of particular interest to this study. Numerical examples are provided to illustrate the differences among the various amplitude estimators. Although amplitude estimation applications are numerous, we focus herein on the problem of system identification using sinusoidal probing signals for which we provide a computationally simple and statistically accurate solution.

Index Terms—Amplitude estimation, spectral analysis, system identification.

I. INTRODUCTION

CONSIDER the noise-corrupted observations of K complex-valued sinusoids

$$x(n) = \sum_{k=1}^K \alpha_k e^{j\omega_k n} + v(n), \quad n = 0, 1, \dots, N-1 \quad (1)$$

where

- α_k complex amplitude of the k th sinusoid having frequency ω_k ;
- N number of available data samples;
- $v(n)$ observation noise, which is complex valued and assumed to be stationary (and possibly colored) with mean zero and finite unknown power spectral density (PSD) $\phi(\omega)$.

Manuscript received August 25, 1998; revised August 4, 1999. This work was supported in part by the Senior Individual Grant Program of the Swedish Foundation for Strategic Research, the National Science Foundation under Grant MIP-9457388, and the Office of Naval Research under Grant N00014-96-0817. The associate editor coordinating the review of this paper and approving it for publication was Prof. P. C. Ching.

P. Stoica is with the Department of Systems and Control, Uppsala University, Uppsala, Sweden (e-mail: ps@syscon.uu.se).

H. Li is with the Department of Electrical and Computer Engineering, University of Florida, Gainesville, FL 32611 USA. He is now with the Department of Electrical and Computer Engineering, Stevens Institute of Technology, Hoboken, NJ 07030 USA (e-mail: hli@stevens-tech.edu).

J. Li is with the Department of Electrical and Computer Engineering, University of Florida, Gainesville, FL 32611 USA (e-mail: li@dsp.ufl.edu).

Publisher Item Identifier S 1053-587X(00)01008-4.

We assume that $\{\omega_k\}_{k=1}^K$ are known, with $\omega_k \neq \omega_l$, for $k \neq l$. The problem of interest is to estimate $\{\alpha_k\}_{k=1}^K$ from the observations $\{x(n)\}_{n=0}^{N-1}$. In this paper, we describe a relatively large number of methods for solving this problem.

Section II discusses least squares (LS) methods, which are widely used for amplitude estimation because they are simple and easy to implement. If we restrict ourselves to estimating only one amplitude at a time, then the LS method reduces to the discrete Fourier transform (DFT) of the data at the frequency of the desired sinusoid, which is computationally more efficient than the LS method that estimates K amplitudes simultaneously. Moreover, estimating one amplitude at a time does not necessarily require exact knowledge of either the number of sinusoids in the data or the frequency location of each sinusoid, which is a desired property in some applications. The disadvantage, however, is that using this one-at-a-time technique gives rather poor amplitude estimates when some sinusoids (that are of interest to us) are close to one another. Statistical analyses that compare the merits of the two LS methods are also provided in Section II.

Since the LS methods completely ignore the correlation of the observation noise, they are, in general, suboptimal. By partitioning the data vector into a number of overlapping subvectors, the covariance matrix of the noise-only part of the data subvectors can be estimated, which makes it possible to use a Markov-like estimator that is optimal in the class of weighted least squares (WLS) techniques [1]. We describe in Section III several ways to estimate the aforementioned covariance matrix that lead to different WLS amplitude estimators. Additionally, we show that if the restriction of estimating one amplitude at a time is again imposed, we obtain two WLS amplitude estimators that are equivalent to the Capon [2], [3] and amplitude and phase estimation (APES) [4] methods extensively used for spectral analysis.

The observation that some general spectral estimators, such as Capon and APES, can be used to solve the problem posed in (1) motivated us to seek other relatively sophisticated spectral analysis techniques that could be used for amplitude estimation. Both Capon and APES belong to the general class of filterbank approaches to spectral estimation [5] that involve splitting the data into possibly overlapping subvectors, passing them through a set of narrowband filters (filterbank) whose center frequencies correspond to those that are of interest to us, and, finally, estimating the spectral density function at those frequencies from the filtered and, hopefully, signal-enhanced data. As we may expect, the key issue of filterbank approaches is the design of the filters. A recent study has suggested the choice of matched filters, which gave rise to the matched-filterbank (MAFI) approach

to spectral estimation [6]. Even though neither Capon nor APES was derived in the MAFI framework (see [2] and [4] for their original derivations), it was found that both are members of the MAFI approach [6]. In the light of the work of [6], we derive in Section IV a generalized MAFI approach to amplitude estimation. Interestingly enough, we show that under certain circumstances, MAFI amplitude estimators have equivalent forms to the WLS methods. However, the MAFI approach is more general than the WLS technique in that the latter is a special case of the former. To show this, a new MAFI amplitude estimator that does not fall into the WLS category is described in Section IV. Other interesting MAFI amplitude estimators may exist, but they have yet to be discovered.

A common feature of the amplitude estimators considered in this paper is that none of them models the observation noise exactly. Even so, all methods are asymptotically statistically efficient, that is, they all achieve the Cramér-Rao bound (CRB) in large samples. However, their finite-sample properties, which are of primary interest to this work, are quite different. Since the finite-sample analysis is intractable in most cases, we use Monte-Carlo simulations in Section V to compare these methods with one another.

The amplitude estimation problem in (1) occurs in a variety of signal processing applications (see, e.g., [7], [8], and the references therein). In Section VI, we discuss its application to system identification. We show that by using sinusoidal probing signals and appropriate amplitude estimators, we can avoid the iterative search required by standard system identification routines, such as the (time-domain) output error method (OEM) [1] and yet achieve CRB-like performance by using a computationally efficient parameter estimator.

In concluding this section, we introduce the following notation to distinguish among the various amplitude estimators. For instance, $\text{LSE}(1, 0, 1)$ denotes the LS estimator that does not split the data (and hence, it uses one data “snapshot”), uses no prefiltering, and estimates one amplitude at a time. Likewise, $\text{MAFI}(L, K, K)$ denotes the MAFI estimator that splits the data into L subvectors, utilizes a bank of K prefilters, and estimates K amplitudes simultaneously. The remaining amplitude estimators are similarly designated.

II. LS AMPLITUDE ESTIMATORS

We consider two LS methods in this section, namely, $\text{LSE}(1, 0, K)$ and $\text{LSE}(1, 0, 1)$.

A. $\text{LSE}(1, 0, K)$

This is perhaps the most direct approach. Let us write the available data sequence in the form

$$\begin{bmatrix} x(0) \\ x(1) \\ \vdots \\ x(N-1) \end{bmatrix} = \begin{bmatrix} 1 & \dots & 1 \\ e^{j\omega_1} & \dots & e^{j\omega_K} \\ \vdots & \vdots & \vdots \\ e^{j(N-1)\omega_1} & \dots & e^{j(N-1)\omega_K} \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_K \end{bmatrix} + \begin{bmatrix} v(0) \\ v(1) \\ \vdots \\ v(N-1) \end{bmatrix} \quad (2)$$

or, with obvious definitions

$$\mathbf{x} = \tilde{\mathbf{A}}\boldsymbol{\alpha} + \mathbf{v} \quad (3)$$

which is a linear regression equation. The LS estimate of $\boldsymbol{\alpha}$ is

$$\hat{\boldsymbol{\alpha}} = (\tilde{\mathbf{A}}^H \tilde{\mathbf{A}})^{-1} \tilde{\mathbf{A}}^H \mathbf{x} \quad (4)$$

where $(\cdot)^H$ denotes the conjugate transpose. Note that the noise is not modeled, even though it may be correlated. Despite this fact, $\text{LSE}(1, 0, K)$ is asymptotically efficient [9]. A relatively simple manner to see this is as follows. First, note that $E\{\hat{\boldsymbol{\alpha}}\} = \boldsymbol{\alpha}$, where $E\{\cdot\}$ denotes the statistical expectation. The mean squared error (MSE) of $\hat{\boldsymbol{\alpha}}$ is

$$\begin{aligned} \text{MSE}\{\hat{\boldsymbol{\alpha}}\} &= \text{cov}\{\hat{\boldsymbol{\alpha}}\} \triangleq E\{(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha})(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha})^H\} \\ &= (\tilde{\mathbf{A}}^H \tilde{\mathbf{A}})^{-1} \tilde{\mathbf{A}}^H \mathbf{W} \tilde{\mathbf{A}} (\tilde{\mathbf{A}}^H \tilde{\mathbf{A}})^{-1} \end{aligned} \quad (5)$$

where $\mathbf{W} \triangleq E\{\mathbf{v}\mathbf{v}^H\}$. Hence, since (see, e.g., [10])

$$\lim_{N \rightarrow \infty} \frac{1}{N} (\tilde{\mathbf{A}}^H \tilde{\mathbf{A}}) = \mathbf{I}_N \quad (6)$$

where \mathbf{I}_N denotes the $N \times N$ identity matrix, and

$$\lim_{N \rightarrow \infty} \frac{1}{N} (\tilde{\mathbf{A}}^H \mathbf{W} \tilde{\mathbf{A}}) = \begin{bmatrix} \phi(\omega_1) & & 0 \\ & \ddots & \\ 0 & & \phi(\omega_K) \end{bmatrix} \quad (7)$$

the asymptotic MSE is given by

$$\lim_{N \rightarrow \infty} N \text{MSE}\{\hat{\boldsymbol{\alpha}}\} = \begin{bmatrix} \phi(\omega_1) & & 0 \\ & \ddots & \\ 0 & & \phi(\omega_K) \end{bmatrix}. \quad (8)$$

Under the mild assumption that $v(n)$ is circularly symmetric Gaussian, the CRB for $\boldsymbol{\alpha}$ is given by (see, e.g., [1])

$$\text{CRB}\{\boldsymbol{\alpha}\} = (\tilde{\mathbf{A}}^H \mathbf{W}^{-1} \tilde{\mathbf{A}})^{-1}. \quad (9)$$

Using the result (see [10] once again)

$$\lim_{N \rightarrow \infty} \frac{1}{N} (\tilde{\mathbf{A}}^H \mathbf{W}^{-1} \tilde{\mathbf{A}}) = \begin{bmatrix} \phi^{-1}(\omega_1) & & 0 \\ & \ddots & \\ 0 & & \phi^{-1}(\omega_K) \end{bmatrix} \quad (10)$$

we obtain

$$\lim_{N \rightarrow \infty} N \text{CRB}\{\boldsymbol{\alpha}\} = \begin{bmatrix} \phi(\omega_1) & & 0 \\ & \ddots & \\ 0 & & \phi(\omega_K) \end{bmatrix} \quad (11)$$

which coincides with (8).

Remark 1: It can be readily checked from (5) and (9) that if $v(n)$ is white, i.e., $\mathbf{W} \sim \mathbf{I}_N$, then $\text{LSE}(1, 0, K)$ is statistically efficient for all $N \geq K$.

B. $\text{LSE}(1, 0, 1)$

Since the observation noise $v(n)$ is not modeled, an idea that reduces the computational burden of the LS approach quite a bit is to include $K - 1$ sinusoids in the noise term and, hence, estimate only one amplitude at a time. In some signal processing

applications, the frequencies $\{\omega_k\}_{k=1}^K$ may be unknown. A typical way to estimate both $\{\alpha_k\}_{k=1}^K$ and $\{\omega_k\}_{k=1}^K$ would consist of estimating just one amplitude for varying frequency and then detecting the peaks in the so-obtained spectrum [5], [7], [8]. As such, the assumption made in Section I that $\{\omega_k\}_{k=1}^K$ are known *a priori* may be relaxed when using the one-at-a-time technique.

There is a somewhat subtle problem with the above technique; the sum of $v(n)$ and $K - 1$ sinusoids no longer has a finite PSD, and hence, one of the previously made assumptions fails. Nevertheless, the idea still works as long as no two sinusoids (that are of interest) are spaced too close to one another, as shown below and later in Section V.

The LSE(1,0,1) is easily derived as

$$\hat{\alpha}_k = \frac{1}{N} \sum_{n=0}^{N-1} x(n) e^{-j\omega_k n}, \quad k = 1, 2, \dots, K \quad (12)$$

which is recognized as the DFT of $\{x(n)\}_{n=0}^{N-1}$ at ω_k . The two estimates in (4) and (12) will be close to one another if $|\omega_k - \omega_l| \gg 2\pi/N$ ($\forall k, l; k \neq l$) [5].

An analysis of LSE(1,0,1) runs as follows. Without loss of generality, let us consider (12) for $k = 1$. The LSE(1,0,1) estimate of α_1 can be written as

$$\hat{\alpha}_1 = (\tilde{\mathbf{a}}^H \tilde{\mathbf{a}})^{-1} \tilde{\mathbf{a}}^H \mathbf{x} \quad (13)$$

where $\tilde{\mathbf{a}} = [1 \ e^{j\omega_1} \ \dots \ e^{j(N-1)\omega_1}]^T$, and where $(\cdot)^T$ denotes the transpose. Taking the expectation of (13) yields

$$E\{\hat{\alpha}_1\} = \alpha_1 + \frac{1}{N} \tilde{\mathbf{a}}^H \bar{\mathbf{A}} \bar{\boldsymbol{\alpha}} \quad (14)$$

where $\bar{\boldsymbol{\alpha}} = [\alpha_2 \ \dots \ \alpha_K]^T$, and $\bar{\mathbf{A}}$ is defined through $[\tilde{\mathbf{a}} \ \bar{\mathbf{A}}] \triangleq \tilde{\mathbf{A}}$. Hence, LSE(1,0,1) is biased. However, it is asymptotically unbiased (that is, its bias goes to zero as $N \rightarrow \infty$). We next calculate the MSE of $\hat{\alpha}_1$ as

$$\text{MSE}\{\hat{\alpha}_1\} = (\tilde{\mathbf{a}}^H \tilde{\mathbf{a}})^{-1} \tilde{\mathbf{a}}^H (\bar{\mathbf{A}} \bar{\boldsymbol{\alpha}} \bar{\boldsymbol{\alpha}}^H \bar{\mathbf{A}}^H + \mathbf{W}) \tilde{\mathbf{a}} (\tilde{\mathbf{a}}^H \tilde{\mathbf{a}})^{-1}. \quad (15)$$

Making use of (7) once again, along with the fact that $\bar{\mathbf{A}}^H \tilde{\mathbf{a}} / \sqrt{N} \rightarrow 0$ as $N \rightarrow \infty$, we have

$$\lim_{N \rightarrow \infty} \text{NMSE}\{\hat{\alpha}_1\} = \phi(\omega_1). \quad (16)$$

Hence, LSE(1,0,1) is also asymptotically efficient. On the other hand, in finite samples, (12) may be better or worse than (4), depending on the characteristics of the scenario under study.

The fact that (4) may be better than (12) comes as no surprise. As an example, let us assume that the signal-to-noise ratio (SNR) is high. Then, the bias of (12) dominates the variance part. On the other hand, (4) has no bias, and its variance will be smaller than the bias of (12) if the SNR is large enough. Consequently, the MSE of (4) will be smaller than that of (12).

The fact that (12) may be better than (4) is, however, a surprise [even though the biased estimator in (12) may *in principle* have a smaller MSE than the CRB for unbiased estimates in (9)]. For

an example of such a case, assume $\text{SNR} \ll 1$ and $\mathbf{W} = \mathbf{I}_N$. Then, for (12)

$$\text{MSE}\{\hat{\alpha}_1\} \approx (\tilde{\mathbf{a}}^H \tilde{\mathbf{a}})^{-1} \quad (17)$$

whereas for (4)

$$\text{MSE}\{\hat{\alpha}_1\} = [(\tilde{\mathbf{A}}^H \tilde{\mathbf{A}})^{-1}]_{1,1} \quad (18)$$

which can be much larger than (17) (e.g., if $|\omega_k - \omega_l| \sim 2\pi/N$ for some $k \geq 2$). In (18), $[\cdot]_{i,j}$ denotes the i th element of the matrix argument.

Note that for most cases of interest, LSE(1,0, K) will give more accurate amplitude estimates than LSE(1,0,1) and that the difference between these two estimators is small for large N . On the other hand, LSE(1,0,1) is computationally more efficient than LSE(1,0, K) since the matrix multiplication and inversion in (4) are avoided. Hence, LSE(1,0,1) may still be worth considering.

III. WLS AMPLITUDE ESTIMATORS

If we partition the data vector \mathbf{x} into (overlapping) subvectors, then the covariance matrix of the noise part of the subvectors may be estimated and can, hence, be used to derive an “optimal” WLS estimator (i.e., a Markov-like estimator) [1]. In this section, we describe a number of such WLS estimators that split the data string into overlapping vectors of shorter length, utilize no prefiltering, and estimate either one or K amplitudes at a time.

A. WLSE($L, 0, K$)

We define the following data subvectors

$$\mathbf{y}(l) = [x(l) \ x(l+1) \ \dots \ x(l+M-1)]^T \quad (19)$$

$$l = 0, 1, \dots, L-1$$

where $L \triangleq N - M + 1$. The choice of M or, equivalently, of L is discussed in Section V. ¹ We have

$$\mathbf{y}(l) = \begin{bmatrix} 1 & \dots & 1 \\ e^{j\omega_1} & \dots & e^{j\omega_K} \\ \vdots & \vdots & \vdots \\ e^{j(M-1)\omega_1} & \dots & e^{j(M-1)\omega_K} \end{bmatrix} \begin{bmatrix} \alpha_1 e^{j\omega_1 l} \\ \alpha_2 e^{j\omega_2 l} \\ \vdots \\ \alpha_K e^{j\omega_K l} \end{bmatrix} + \begin{bmatrix} v(l) \\ v(l+1) \\ \vdots \\ v(l+M-1) \end{bmatrix} \quad (20)$$

or, with obvious notation

$$\mathbf{y}(l) = \mathbf{A} \mathbf{s}(l) + \boldsymbol{\epsilon}(l). \quad (21)$$

Alternatively, we can rewrite (20) as

$$\mathbf{y}(l) = \mathbf{A}_l \boldsymbol{\alpha} + \boldsymbol{\epsilon}(l) \quad (22)$$

¹ M may be chosen smaller than K , see, e.g., Fig. 4. Moreover, when $M = 1$, all WLSE($L, 0, K$) reduce to LSE(1,0, K).

where

$$\mathbf{A}_l \triangleq \mathbf{A} \begin{bmatrix} e^{j\omega_1 l} & & 0 \\ & \ddots & \\ 0 & & e^{j\omega_K l} \end{bmatrix} \triangleq \mathbf{A} \mathbf{D}_l. \quad (23)$$

We will use (21) mostly for analysis and (22) for estimation.

The WLS (Markov-like) estimate of $\boldsymbol{\alpha}$ in (22) is given by

$$\hat{\boldsymbol{\alpha}} = \left[\sum_{l=0}^{L-1} \mathbf{A}_l^H \hat{\mathbf{Q}}^{-1} \mathbf{A}_l \right]^{-1} \left[\sum_{l=0}^{L-1} \mathbf{A}_l^H \hat{\mathbf{Q}}^{-1} \mathbf{y}(l) \right] \quad (24)$$

where $\hat{\mathbf{Q}}$ is an estimate of $\mathbf{Q} = E\{\boldsymbol{\epsilon}(l)\boldsymbol{\epsilon}^H(l)\}$.

To estimate \mathbf{Q} , we may proceed as follows. Let

$$\hat{\mathbf{R}} = \frac{1}{L} \sum_{l=0}^{L-1} \mathbf{y}(l)\mathbf{y}^H(l). \quad (25)$$

We can verify that as $L \rightarrow \infty$, $\hat{\mathbf{R}}$ goes to

$$\mathbf{R} = \mathbf{A} \mathbf{P} \mathbf{A}^H + \mathbf{Q} \quad (26)$$

where

$$\mathbf{P} = \begin{bmatrix} |\alpha_1|^2 & & 0 \\ & \ddots & \\ 0 & & |\alpha_K|^2 \end{bmatrix}. \quad (27)$$

Hence, one way to estimate \mathbf{Q} is as

$$\hat{\mathbf{Q}} = \hat{\mathbf{R}} - \hat{\mathbf{A}} \hat{\mathbf{P}} \hat{\mathbf{A}}^H \quad (28)$$

where $\hat{\mathbf{P}}$ is made from some initial estimates of $\{\alpha_k\}_{k=1}^K$ obtained for instance via one of the LS amplitude estimators. The need for initial amplitude estimates is a drawback of $\hat{\mathbf{Q}}$ in (28). In the following, we try to circumvent this need in two different ways.

First, we show a way to simplify the WLSE($L, 0, K$) that uses (24) with (28). From (28), we have that

$$\hat{\mathbf{R}} \hat{\mathbf{Q}}^{-1} \mathbf{A} = \hat{\mathbf{A}} \hat{\mathbf{P}} \hat{\mathbf{A}}^H \hat{\mathbf{Q}}^{-1} \mathbf{A} + \mathbf{A} = \mathbf{A} \boldsymbol{\Gamma} \quad (29)$$

where

$$\boldsymbol{\Gamma} \triangleq \hat{\mathbf{P}} \hat{\mathbf{A}}^H \hat{\mathbf{Q}}^{-1} \mathbf{A} + \mathbf{I}_K. \quad (30)$$

For sufficiently large N and M , $\boldsymbol{\Gamma}$ is approximately diagonal since $\mathbf{A}^H \hat{\mathbf{Q}}^{-1} \mathbf{A}$ is so (see, e.g., (10)). Consequently

$$\hat{\mathbf{Q}}^{-1} \mathbf{A}_l = \hat{\mathbf{R}}^{-1} \mathbf{A} \boldsymbol{\Gamma} \mathbf{D}_l \approx \hat{\mathbf{R}}^{-1} \mathbf{A} \mathbf{D}_l \boldsymbol{\Gamma} = \hat{\mathbf{R}}^{-1} \mathbf{A}_l \boldsymbol{\Gamma}. \quad (31)$$

Inserting (31) into (24) yields (observe that $\boldsymbol{\Gamma}^H$ cancels out)

$$\hat{\boldsymbol{\alpha}} \approx \left[\sum_{l=0}^{L-1} \mathbf{A}_l^H \hat{\mathbf{R}}^{-1} \mathbf{A}_l \right]^{-1} \left[\sum_{l=0}^{L-1} \mathbf{A}_l^H \hat{\mathbf{R}}^{-1} \mathbf{y}(l) \right] \quad (32)$$

which, unlike using (24) with (28), does not require any initial estimate of $\{\alpha_k\}_{k=1}^K$. The amplitude estimator in (32) can be interpreted as an *extension of the Capon algorithm* in [2] and [3] to multiple sinusoids.

A different estimate of \mathbf{Q} can be obtained as described next. Observe that

$$\mathbf{A} \mathbf{P} \mathbf{A}^H = \sum_{k=1}^K [\alpha_k \mathbf{a}(\omega_k)] [\alpha_k \mathbf{a}(\omega_k)]^H \triangleq \sum_{k=1}^K \boldsymbol{\beta}_k \boldsymbol{\beta}_k^H \quad (33)$$

where

$$\mathbf{a}(\omega) = [1 \quad e^{j\omega} \quad \dots \quad e^{j(M-1)\omega}]^T. \quad (34)$$

We can use the vectors $\{\boldsymbol{\beta}_k\}_{k=1}^K$ introduced above to rewrite (22) as

$$\mathbf{y}(l) = \sum_{k=1}^K \boldsymbol{\beta}_k e^{j\omega_k l} + \boldsymbol{\epsilon}(l). \quad (35)$$

From (35), we can estimate $\boldsymbol{\beta}_k$ one at a time via LS as

$$\hat{\boldsymbol{\beta}}_k = \frac{1}{L} \sum_{l=0}^{L-1} \mathbf{y}(l) e^{-j\omega_k l} \triangleq \mathbf{g}(\omega_k), \quad k = 1, 2, \dots, K. \quad (36)$$

(Note that we could estimate all $\{\boldsymbol{\beta}_k\}_{k=1}^K$ simultaneously via LS, which is, however, more complicated and does not appear to perform better than (36).) The use of (36) in (26) and (33) leads to the following estimate of \mathbf{Q} :

$$\hat{\mathbf{Q}} = \hat{\mathbf{R}} - \sum_{k=1}^K \mathbf{g}(\omega_k) \mathbf{g}^H(\omega_k). \quad (37)$$

The estimate of \mathbf{Q} in (37) is different from the one in (28), yet we use the same symbol for both of them to simplify the notation. The WLSE($L, 0, K$) that uses (24) with (37) does not require any initial estimate of $\{\alpha_k\}_{k=1}^K$. It is an *extension of the APES algorithm* in [4] to multiple sinusoids with known frequencies.

Remark 2: We note that $\boldsymbol{\epsilon}(k)$ and $\boldsymbol{\epsilon}(l)$ in (22) are correlated (for $k \neq l$), which implies that (24) is suboptimal (as it takes into account only the correlation between the elements of $\boldsymbol{\epsilon}(l)$ but ignores the correlation between $\boldsymbol{\epsilon}(l)$ and $\boldsymbol{\epsilon}(k)$ for $k \neq l$). Yet, the WLS methods are likely to outperform the LS methods because the latter completely ignore the correlation in $v(n)$.

B. WLSE($L, 0, 1$)

The particularization of WLSE($L, 0, K$) to WLSE($L, 0, 1$) is straightforward. Specifically, the WLSE($L, 0, 1$) that corresponds to using (24) with (28) can be readily verified (by using the matrix inversion lemma) to be

$$\hat{\alpha}_k = \frac{\mathbf{a}^H(\omega_k) \hat{\mathbf{R}}^{-1} \mathbf{g}(\omega_k)}{\mathbf{a}^H(\omega_k) \hat{\mathbf{R}}^{-1} \mathbf{a}(\omega_k)}, \quad k = 1, 2, \dots, K \quad (38)$$

whereas the WLSE($L, 0, 1$) that corresponds to using (24) with (37) is given by

$$\hat{\alpha}_k = \frac{\mathbf{a}^H(\omega_k) [\hat{\mathbf{R}} - \mathbf{g}(\omega_k) \mathbf{g}^H(\omega_k)]^{-1} \mathbf{g}(\omega_k)}{\mathbf{a}^H(\omega_k) [\hat{\mathbf{R}} - \mathbf{g}(\omega_k) \mathbf{g}^H(\omega_k)]^{-1} \mathbf{a}(\omega_k)}, \quad k = 1, 2, \dots, K. \quad (39)$$

Note that (38), like (32), does not depend on $\hat{\mathbf{P}}$. However, unlike (32), (38) is *exactly* equivalent to using (24) with (28). Equations (38) and (39) have the same form as the Capon [2], [3] and, respectively, the APES [4] spectral estimators. The latter

two estimators were derived in [6] and [11] by a different approach, namely, the MAFI approach, which we will consider in a generalized form in the next section.

It is interesting that the above two amplitude estimators, while both *asymptotically efficient* (and, hence, equivalent), have quite *different finite-sample properties*. In particular, it was shown in [6] and [11] that (38) is biased downward, whereas (39) is unbiased (within a second-order approximation) and, in general, has a better performance than the former.

IV. MAFI AMPLITUDE ESTIMATORS

In this section, we derive a generalized MAFI approach to amplitude estimation. Let $\mathbf{H}^H \in \mathbb{C}^{\bar{K} \times M}$ be a matrix, each row of which is a finite impulse response (FIR) filter (for some $1 \leq \bar{K} \leq M$ yet to be specified). The MAFI idea can be explained as follows.

- Design \mathbf{H}^H so that when applied to $\{\mathbf{y}(l)\}$, it maximizes the SNR at the \bar{K} filter outputs.
- Estimate the amplitudes from the filtered data (whose SNR should be higher than that in the raw data) by, e.g., the LS or WLS technique.

Mathematically, \mathbf{H} can be obtained as

$$\mathbf{H} = \arg \max_{\mathbf{H}} \underbrace{\text{tr}[(\mathbf{H}^H \hat{\mathbf{Q}} \mathbf{H})^{-1} \mathbf{H}^H (\mathbf{A} \hat{\mathbf{P}} \mathbf{A}^H) \mathbf{H}]}_{\text{"Generalized SNR"}} \quad (40)$$

where \mathbf{H} is constrained in a way that is specified later, and $\text{tr}(\cdot)$ denotes the trace of a matrix. Let

$$\mathbf{X}^H = (\mathbf{H}^H \hat{\mathbf{Q}} \mathbf{H})^{-1/2} \mathbf{H}^H \hat{\mathbf{Q}}^{1/2} \quad (41)$$

where $(\cdot)^{1/2}$ denotes the Hermitian square root of the positive definite matrix argument. Observe that \mathbf{X} is semiunitary, i.e., $\mathbf{X}^H \mathbf{X} = \mathbf{I}_{\bar{K}}$. The cost function in (40) can now be rewritten as

$$f = \text{tr}[\mathbf{X}^H \hat{\mathbf{Q}}^{-1/2} \mathbf{A} \hat{\mathbf{P}} \mathbf{A}^H \hat{\mathbf{Q}}^{-1/2} \mathbf{X}]. \quad (42)$$

It follows from the Poincaré separation theorem (or the generalized Rayleigh quotient theorem) [12] that

$$\max_{\mathbf{X}} f = \sum_{k=1}^{\bar{K}} \lambda_k(\hat{\mathbf{Q}}^{-1/2} \mathbf{A} \hat{\mathbf{P}} \mathbf{A}^H \hat{\mathbf{Q}}^{-1/2}) \quad (43)$$

where $\{\lambda_k(\cdot)\}$ denotes the eigenvalues of the matrix between the parentheses, ordered such that $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_M$; furthermore, the columns of the maximizing \mathbf{X} are equal to the eigenvectors corresponding to $\{\lambda_k\}_{k=1}^{\bar{K}}$.

Next, note that postmultiplying \mathbf{X} by any unitary matrix of appropriate dimensions yields another valid solution for \mathbf{X} . One such solution having a simple form can be obtained as follows. Observe that

$$\text{rank}(\hat{\mathbf{Q}}^{-1/2} \mathbf{A} \hat{\mathbf{P}} \mathbf{A}^H \hat{\mathbf{Q}}^{-1/2}) = K \quad (44)$$

which implies that we cannot improve the generalized SNR by choosing $\bar{K} > K$ since $\lambda_{K+1} = \dots = \lambda_M = 0$. On the other hand, the larger the $\bar{K} \leq K$, the higher the optimal SNR in (43) and the more filtered data will be available for amplitude estimation. Hence, we choose $\bar{K} = K$. In such a case, maximizing

\mathbf{X} has the same range space as $\hat{\mathbf{Q}}^{-1/2} \mathbf{A}$, and hence, it is given by

$$\mathbf{X} = \hat{\mathbf{Q}}^{-1/2} \mathbf{A} \mathbf{T} \quad (45)$$

where \mathbf{T} denotes some nonsingular matrix that makes \mathbf{X} semiunitary. One such \mathbf{T} is

$$\mathbf{T} = (\mathbf{A}^H \hat{\mathbf{Q}}^{-1} \mathbf{A})^{-1/2}. \quad (46)$$

Hence

$$\mathbf{X} = \hat{\mathbf{Q}}^{-1/2} \mathbf{A} (\mathbf{A}^H \hat{\mathbf{Q}}^{-1} \mathbf{A})^{-1/2}. \quad (47)$$

We next observe that

$$\mathbf{H} = \hat{\mathbf{Q}}^{-1/2} \mathbf{X} \quad (48)$$

satisfies (41). Consequently, we have

$$\mathbf{H} = \hat{\mathbf{Q}}^{-1} \mathbf{A} (\mathbf{A}^H \hat{\mathbf{Q}}^{-1} \mathbf{A})^{-1/2}. \quad (49)$$

The final step is to observe that postmultiplying \mathbf{H} by a nonsingular matrix does not change the generalized SNR criterion. Then, it follows immediately that

$$\mathbf{H} = \hat{\mathbf{Q}}^{-1} \mathbf{A} (\mathbf{A}^H \hat{\mathbf{Q}}^{-1} \mathbf{A})^{-1} \quad (50)$$

maximizes the generalized SNR and it also satisfies the constraint

$$\mathbf{H}^H \mathbf{A} = \mathbf{I}_K. \quad (51)$$

The constraint (51) says that each (row) filter in \mathbf{H}^H passes one sinusoid undistorted and completely annihilates the others.

From (22) and (23), the filtered data corresponding to (50) is given by

$$\begin{aligned} \mathbf{z}(l) &\triangleq \mathbf{H}^H \mathbf{y}(l) = \mathbf{D}_l \boldsymbol{\alpha} + \mathbf{H}^H \boldsymbol{\epsilon}(l) \triangleq \mathbf{D}_l \boldsymbol{\alpha} + \boldsymbol{\nu}(l) \\ &l = 0, 1, \dots, L-1. \end{aligned} \quad (52)$$

The covariance matrix of $\boldsymbol{\nu}(l)$ can be estimated as

$$\mathbf{H}^H \hat{\mathbf{Q}} \mathbf{H} = (\mathbf{A}^H \hat{\mathbf{Q}}^{-1} \mathbf{A})^{-1}. \quad (53)$$

It follows that the WLS (Markov-like) estimate of $\boldsymbol{\alpha}$ in (52) is given by (54), shown at the bottom of the next page, which shows that

$$\text{MAFI}(L, K, K) = \text{WLSE}(L, 0, K). \quad (55)$$

The MAFI interpretation of the WLS method, which is afforded by the above analysis, is interesting. In particular, it makes a clear connection between using the MAFI and the WLS techniques for amplitude estimation. The MAFI approach is, however, more general than the WLS technique. As an example, we derive a new MAFI amplitude estimator that does not belong to the WLS class. Let $z_k(l)$ and $\nu_k(l)$ denote the k th element of $\mathbf{z}(l)$ and, respectively, $\boldsymbol{\nu}(l)$ in (52). Then

$$z_k(l) = \alpha_k e^{j\omega_k l} + \nu_k(l), \quad k = 1, 2, \dots, K. \quad (56)$$

The above equations are related to one another *only* via the correlation between $\nu_k(l)$ and $\nu_p(l)$ (for $k \neq p$). If we ignore

that correlation, then we can apply LS to (56) to obtain the $\text{MAFI}(L, K, 1)$ estimate of α_k given by

$$\hat{\alpha}_k = \frac{1}{L} \sum_{l=0}^{L-1} z_k(l) e^{-j\omega_k l}. \quad (57)$$

Unlike the Capon (38) and APES (39) estimators [which can also be shown to be members of the $\text{MAFI}(L, K, 1)$ class [6]], the above $\text{MAFI}(L, K, 1)$ estimator *does* require the knowledge of the number and frequencies of the sinusoids, which makes it behave more like a $\text{MAFI}(L, K, K)$ estimator. In particular, it performs quite well for cases where some sinusoids are closely spaced, as will be seen in Section V.

Other interesting MAFI amplitude estimators may be devised by using some other choices of \mathbf{H} in lieu of the one given in (50), as the solution to (40) is not unique. Specifically, we may use a linear transformation on the \mathbf{H} in (50), or choose $\bar{K} < K$, or replace the \mathbf{I}_K in (51) by another nonsingular matrix, all of these modifications leading to solutions that are different from (50). Furthermore, we could even change the criterion in (40) to another reasonable definition of the ‘‘generalized SNR.’’ However, such variations on the theme of MAFI are beyond the scope of the present paper.

V. NUMERICAL EXAMPLES

In what follows, we investigate the performances of the various amplitude estimators described in the previous sections. For notational simplicity, we will refer to these methods as follows:

- LSE1: LSE(1, 0, 1) using (12);
- LSEK: LSE(1, 0, K) using (4);
- Capon1: WLSE(L , 0, 1) using (38);
- APES1: WLSE(L , 0, 1) using (39);
- CaponK: WLSE(L , 0, K) using (32);
- APESK: WLSE(L , 0, K) using (24) along with (37);
- MAFI1: $\text{MAFI}(L, K, 1)$ using (57) along with (37).

We will compare these methods with one another as well as the CRB given in (9). Since all these methods are asymptotically efficient, we only consider the case when N is relatively small. Specifically, we choose $N = 32$. The data consist of three complex sinusoids corrupted by a complex Gaussian noise $v(n)$ (to be specified)

$$x(n) = s(n) + v(n), \quad n = 0, 1, \dots, N - 1 \quad (58)$$

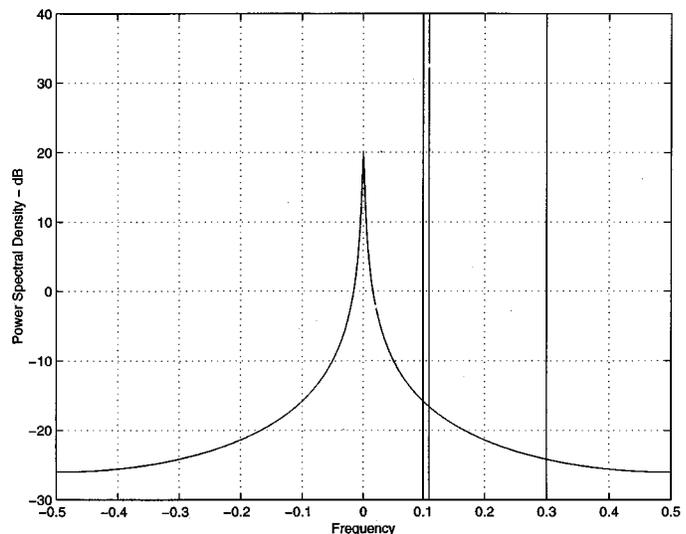


Fig. 1. PSD of the test data that consist of three sinusoids and an AR(1) noise process.

where

$$s(n) = \sum_{k=1}^3 \alpha_k e^{j2\pi f_k n}. \quad (59)$$

The frequencies of the sinusoids are $f_1 = 0.1$, $f_2 = 0.11$, and $f_3 = 0.3$. In addition, $\alpha_1 = e^{j\pi/4}$, $\alpha_2 = e^{j\pi/3}$, and $\alpha_3 = e^{j\pi/4}$.

All examples are based on 200 Monte Carlo simulations. The MSE figures shown in what follows are obtained as

$$\text{MSE}\{\hat{\alpha}_k\} = \frac{1}{200} \sum_{i=1}^{200} |\hat{\alpha}_k(i) - \alpha_k|^2 \quad (60)$$

where $\hat{\alpha}_k(i)$ is the estimate of α_k derived in the i th simulation run.

A. Estimation Performance versus SNR

First, we consider the case where $v(n)$ is colored. More exactly, $v(n)$ is described by the following autoregressive (AR) equation

$$v(n) = 0.99v(n-1) + e(n) \quad (61)$$

where $e(n)$ is a complex white Gaussian noise with zero-mean and variance σ^2 . The PSD of the test data is shown in Fig. 1,

$$\begin{aligned} \hat{\alpha} &= \left[\sum_{l=0}^{L-1} \mathbf{D}_l^H (\mathbf{A}^H \hat{\mathbf{Q}}^{-1} \mathbf{A}) \mathbf{D}_l \right]^{-1} \left[\sum_{l=0}^{L-1} \mathbf{D}_l^H (\mathbf{A}^H \hat{\mathbf{Q}}^{-1} \mathbf{A}) (\mathbf{A}^H \hat{\mathbf{Q}}^{-1} \mathbf{A})^{-1} \mathbf{A}^H \hat{\mathbf{Q}}^{-1} \mathbf{y}(l) \right] \\ &= \left[\sum_{l=0}^{L-1} \mathbf{A}_l^H \hat{\mathbf{Q}}^{-1} \mathbf{A}_l \right]^{-1} \left[\sum_{l=0}^{L-1} \mathbf{A}_l^H \hat{\mathbf{Q}}^{-1} \mathbf{y}(l) \right] \end{aligned} \quad (54)$$

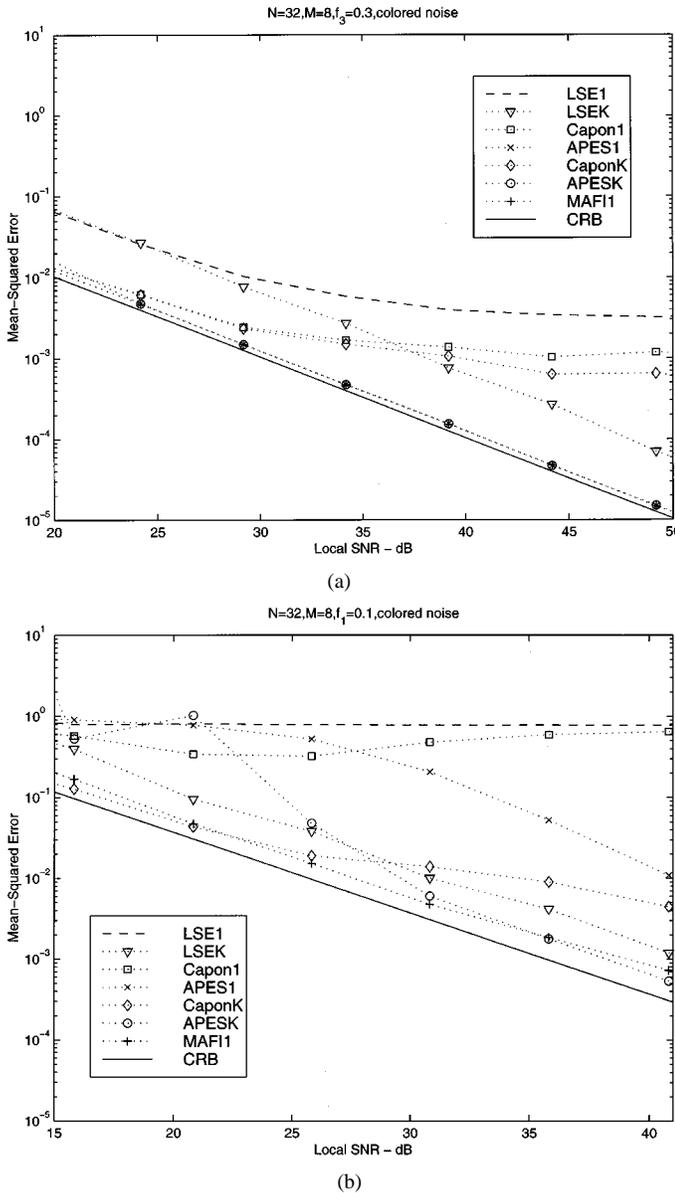


Fig. 2. Empirical MSE's and the CRB versus local SNR when $N = 32$, $M = 8$, and the observation noise is colored [an AR(1) process]. (a) α_3 . (b) α_1 .

where $\sigma^2 = 0.01$. The *local* SNR of the k th sinusoid is defined as [7]

$$\text{SNR}_k = 10 \log_{10} \frac{|\alpha_k|^2}{\int_{f_k-1/(2N)}^{f_k+1/(2N)} \phi(f) df} \approx 10 \log_{10} \frac{N|\alpha_k|^2}{\phi(f_k)}. \quad (62)$$

Note the occurrence of N in the above SNR formula. For those methods that depend on M , we choose $M = N/4 = 8$, giving $L = 25$ (see Section V-B for a study of the effect of M on the performance).

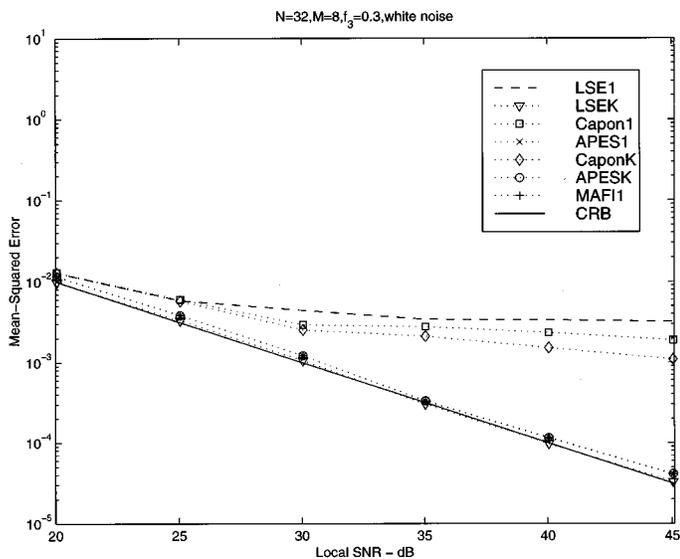
Fig. 2(a) shows the MSE's of the seven amplitude estimators for α_3 along with the corresponding CRB as the SNR varies. As we can see, APES1, APESK, and MAFI1 are very close to the CRB, whereas LSEK, which ignores the noise correlation, is evidently away from the CRB. CaponK also deviates from

the CRB for most SNR's. The reason is that the approximation made in (31) is valid only for large N and M , which is not the case in this example. Fig. 2(a) also shows that both LSE1 and Capon1 are inconsistent (in SNR). Their inconsistency is not surprising because both are biased estimators. Recall that the bias of LSE1, as given in (14), does not vanish unless N goes to infinity. Similarly, Capon1 is always biased (downward) for finite N [6], [11].

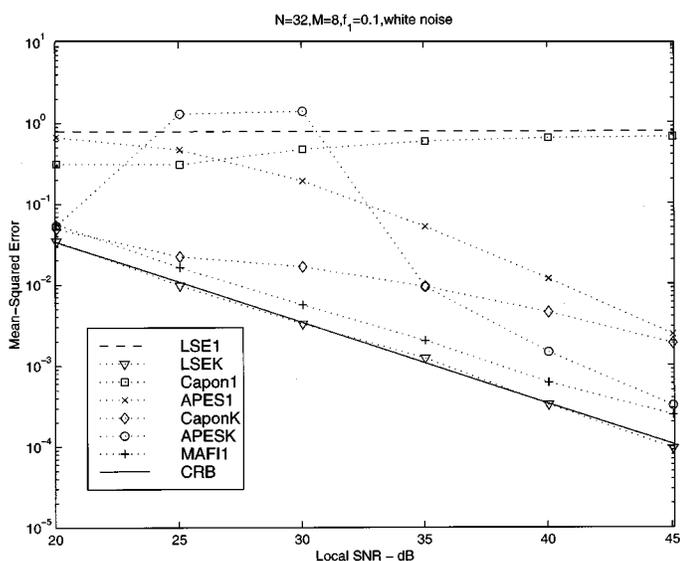
Fig. 2(b) shows the counterpart curves for α_1 . (The results for α_2 are omitted because they resemble those for α_1 .) Note that $f_2 - f_1 = 0.01$, which is smaller than $1/N \approx 0.03$ (the Fourier resolution limit). The performance degrades for all estimators under study, especially for LSE1, Capon1, and APES1, which estimate only one amplitude at a time. As shown in Fig. 2(b), LSE1 and Capon1 essentially fail for all SNR's considered due to their large MSE's. APES1 is no longer close to the CRB but, unlike the previous two estimators, it still appears to be consistent (in SNR). As in Fig. 2(a), CaponK again deviates away from the CRB at high SNR's. It appears that the approximation made in (31) introduces a bias (at small N and M) that may be negligible at low SNR's but dominates the variance at high SNR's. The bias does not disappear as the SNR increases, which causes the divergence of CaponK from the CRB. APESK performs quite well for high SNR's; however, it is not very stable at low SNR's (due to large variance). The best estimator in this example is MAFI1. The knowledge of the number and locations of the sinusoids, which the other one-at-a-time estimators do not necessarily need but is required by MAFI1, appears to play an important role in the current case.

As stated in Section II, LSEK is statistically efficient, i.e., it achieves the CRB for any $N \geq K$ when the observation noise is white. To see how the other suboptimal (in finite samples) methods perform in such a case, we consider an example that is similar to the previous one, except that $v(n)$ is replaced by a zero-mean complex white Gaussian noise. The SNR is defined in the same manner as in (62). Fig. 3(a) and (b) show the MSE's of the amplitude estimates of α_3 and, respectively, α_1 and the corresponding CRB as the SNR increases. As we can see, the APES1, APESK, and MAFI1 estimates of α_3 are again very close to the CRB, whereas for α_1 , all suboptimal methods suffer from some performance loss, as compared with the optimal LSEK, and yet the differences between LSEK and MAFI1 for all SNR's considered here are fairly small.

A brief summary based on the previous study is as follows. *APES1 is recommended in applications where it is known a priori that no two sinusoids are closely spaced* (see, e.g., the application discussed in the next section) or when the closely spaced sinusoids are of no interest. The reason to prefer APES1 to APESK or MAFI1 in such cases is that the former is more flexible than the latter two since APES1 does not necessarily require knowledge of the sinusoidal frequencies. In terms of computational cost, APES1 and MAFI1 are similar to one another, and both are simpler than APESK. When it is desired to estimate the amplitudes of closely spaced sinusoids in colored noise, however, MAFI1 may be preferred. In general, we do not recommend the use of Capon1 since it has a computational complexity similar to that of APES1, but it is biased. Although we did notice that CaponK gives close-to-CRB performance at very



(a)



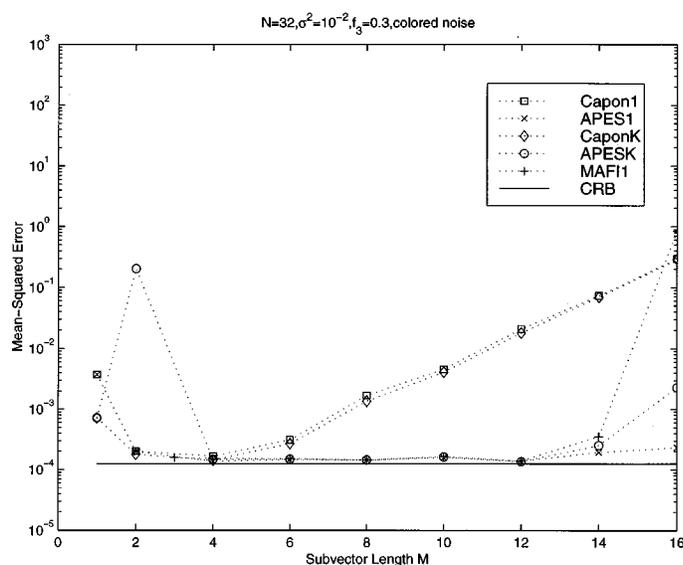
(b)

Fig. 3. Empirical MSE's and the CRB versus local SNR when $N = 32$, $M = 8$, and the observation noise is white. (a) α_3 . (b) α_1 .

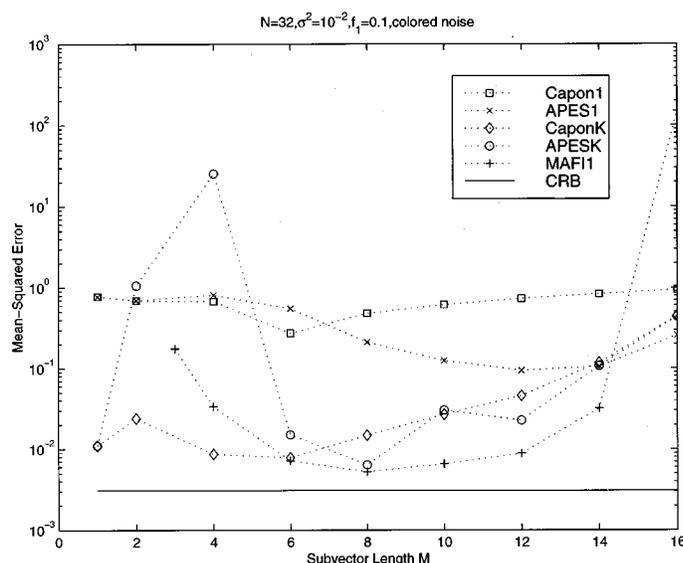
low SNR's, in most cases of interest, other methods like APES1 or MAFI1 may be preferred. *LSEK is statistically efficient and may be preferred when the observation noise is white*; in cases where the white noise assumption is invalid, it is preferable to use APES1 or MAFI1. LSE1 gives comparatively rather poor estimation accuracy but is computationally quite simple. The performance differences stated so far occur only when N is relatively small. As N increases, the performance of all methods tends to the CRB, independently of the noise correlation. Hence, *when N is sufficiently large, LSE1 should be preferred because of its computational simplicity.*

B. The Effect of M

All WLS and MAFI amplitude estimators studied in this paper depend on the choice of M : the subvector length. It is known that as M increases, all of them can better deal with the case of closely spaced sinusoids, but their statistical stability, in



(a)



(b)

Fig. 4. Empirical MSE's and the CRB versus M when $N = 32$ and the observation noise is colored [an AR(1) process with $\sigma^2 = 0.01$]. (a) α_3 . (b) α_1 .

general, decreases [5]. Hence, there is a tradeoff to be kept in mind when choosing M . Note that M should also be smaller than $N/2$; otherwise, the estimated covariance matrix will be rank deficient. The following example examines the effect of M on the performances of these estimators. LSE1 and LSEK do not depend on M and are thus not considered in this example.

The scenario is similar to the first example (AR noise), except that we fix $\sigma^2 = 10^{-2}$, which corresponds to a local SNR of 30.8 dB for the first sinusoid (at $f_1 = 0.1$) and 39.2 dB for the third sinusoid (at $f_3 = 0.3$). M is varied from 1 to 16 for all estimators, except for MAFI1, which requires that $M \geq K$ [see (37) and (50)]. The MSE's of the amplitude estimates of α_3 and, respectively, α_1 , and the corresponding CRB's are shown in Fig. 4(a) and (b). As can be seen from the figure, all estimators are sensitive to the choice of M to a smaller or larger extent. When no sinusoids are close to the one being estimated, such as

TABLE I
CHOICE OF M FOR THE WLS AND MAFI
AMPLITUDE ESTIMATORS

Estimator	M
APES1	$N/4 \leq M \leq N/2$
APESK	$N/4 \leq M \leq N/3$
MAFI1	$N/8 \leq M \leq 2N/5$
Capon1 or CaponK	$N/8 \leq M \leq N/4$

the third sinusoid in this example, APES1, APESK, and MAFI1 perform quite well for a wide range of M . For the more difficult case shown in Fig. 4(b), the choice of M becomes critical. Based on our empirical observations, some rules of thumb for the choice of M are given in Table I.

C. Imprecise Knowledge of the Frequencies

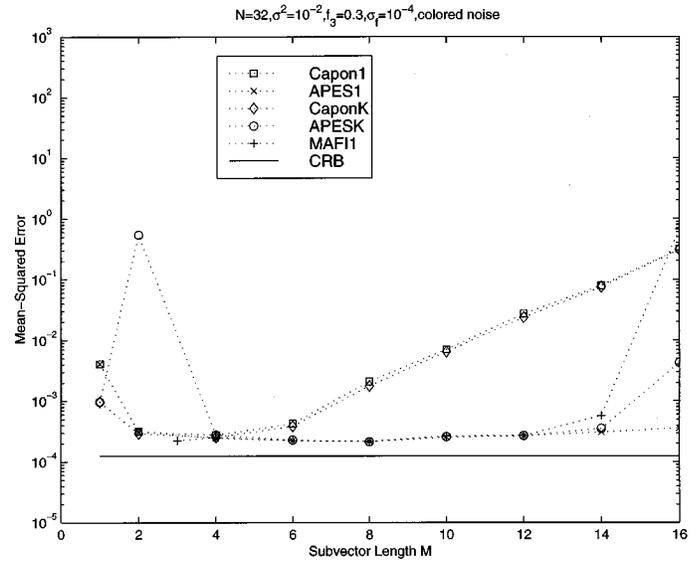
The emphasis in this paper has been on amplitude estimation. The estimation of the frequencies was briefly mentioned but is not in the main focus of the paper. However, in applications, the frequencies are rarely “exactly” known. Consequently, it is an interesting question as to whether the conclusions about and the assessment of the performance of various amplitude estimation methods remain essentially the same when the frequencies are imprecisely known. To answer this question, we repeated the experiment in Section V-B under the conditions described there, with one exception; in each realization, the frequencies used by the amplitude estimators were generated by adding to the true values a set of independent and identically distributed Gaussian random variables with zero mean and standard deviation denoted by σ_f .

Figs. 5 and 6 show the results obtained for $\sigma_f = 10^{-4}$ and $\sigma_f = 10^{-3}$, respectively. The CRB plots in the figures are the same as those in Fig. 4 corresponding to the case when the frequencies were exactly known ($\sigma_f = 0$). Comparing the results in Figs. 4 and 5, we can see that they do not differ much from one another. To gain an appreciation of this observation, we note that high-quality oscillators have frequency errors of an order typically smaller than $\sigma_f = 10^{-4}$. For the case of the larger frequency errors considered corresponding to $\sigma_f = 10^{-3}$, all amplitude estimators under discussion suffer a substantial performance degradation (compare Figs. 4 and 6), especially APESK, whose performance is apparently affected by outliers and is, hence, relatively unstable (even in the case of $\sigma_f = 0$; see Fig. 4). However, the ranking of the estimators and the way their performance varies with M in the case of $\sigma_f = 10^{-3}$ are similar to those observed in the case of $\sigma_f = 0$.

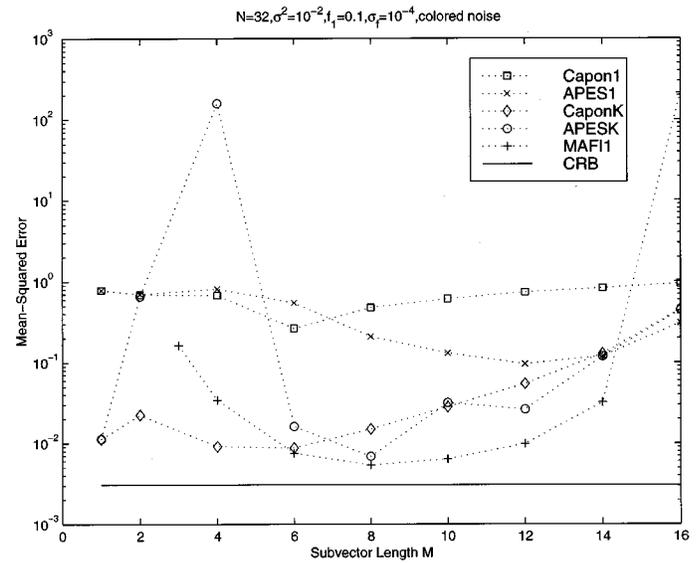
VI. APPLICATION TO SYSTEM IDENTIFICATION

Consider the linear discrete-time system described by [1]

$$x(n) = H(z^{-1})u(n) + v(n), \quad n = 0, 1, \dots, N-1 \quad (63)$$



(a)



(b)

Fig. 5. Empirical MSE's and the CRB versus M when $N = 32$, $\sigma_f = 10^{-4}$, and the observation noise is colored [an AR(1) process with $\sigma^2 = 0.01$]. (a) α_3 . (b) α_1 .

where the input $u(n)$ is a sinusoidal (probing) signal

$$u(n) = \sum_{k=1}^K \gamma_k e^{j\omega_k n}, \quad n = 0, 1, \dots, N-1 \quad (64)$$

and the transfer function is rational

$$H(z^{-1}) = \frac{B(z^{-1})}{A(z^{-1})} = \frac{b_1 z^{-1} + \dots + b_q z^{-q}}{1 + a_1 z^{-1} + \dots + a_p z^{-p}}. \quad (65)$$

We assume that

$$K \geq p + q. \quad (66)$$

Even if p and q were unknown, K could still be chosen sufficiently large to satisfy (66). The problem of interest in this section is to estimate $\{a_i\}_{i=1}^p$ and $\{b_j\}_{j=1}^q$ from $\{x(n)\}_{n=0}^{N-1}$.

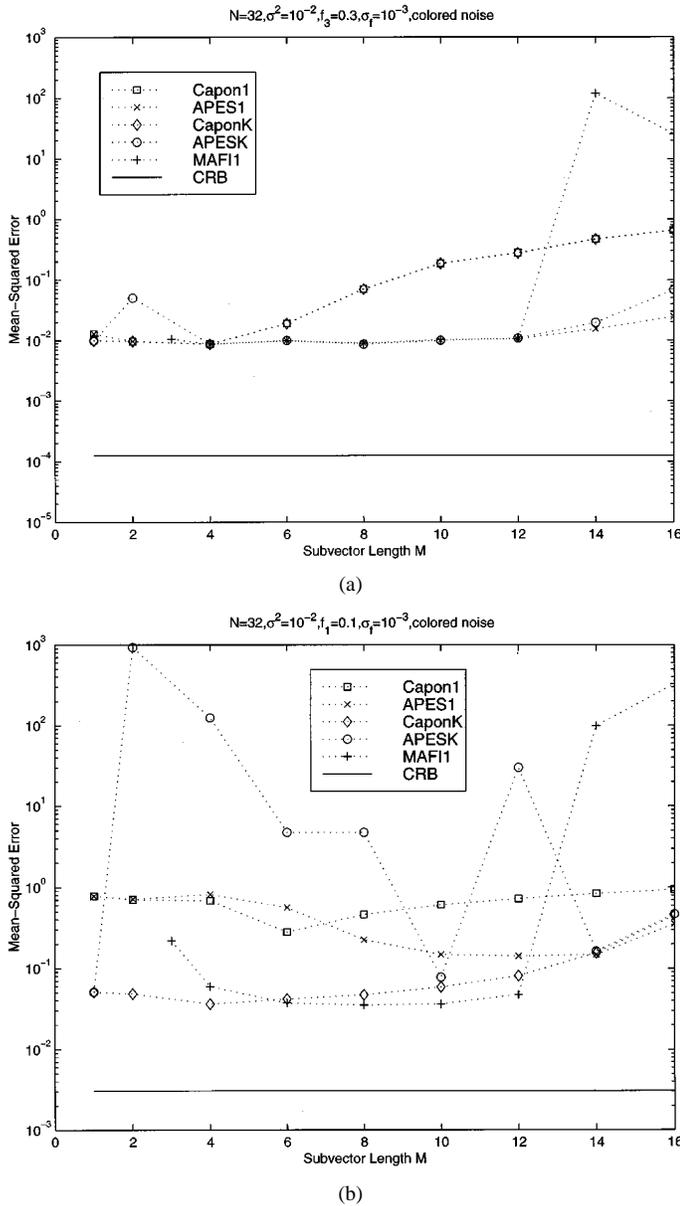


Fig. 6. Empirical MSE's and the CRB versus M when $N = 32$, $\sigma_f = 10^{-3}$, and the observation noise is colored [an AR(1) process with $\sigma^2 = 0.01$]. (a) α_3 . (b) α_1 .

A. System Identification Using Amplitude Estimation

The system identification problem associated with (63) is commonly solved by using the output error method (OEM), which does not model $v(n)$, and obtains estimates of $\{a_i\}_{i=1}^p$ and $\{b_j\}_{j=1}^q$ by minimizing the criterion

$$C_{\text{OEM}}(\mathbf{a}, \mathbf{b}) = \sum_{n=0}^{N-1} |x(n) - H(z^{-1})u(n)|^2 \quad (67)$$

where $\mathbf{a} = [a_1 \cdots a_p]^T$, and $\mathbf{b} = [b_1 \cdots b_q]^T$. OEM is a natural choice of a method whose performance should be compared with that of the new method proposed in this paper.

To describe the latter method, let

$$\alpha_k(\mathbf{a}, \mathbf{b}) = \gamma_k H(e^{-j\omega_k}). \quad (68)$$

For sufficiently large N (so that the transient response in the output can be neglected), the system description in (63) and (64) is approximately equivalent to

$$x(n) = \sum_{k=1}^K \alpha_k(\mathbf{a}, \mathbf{b}) e^{j\omega_k n} + v(n). \quad (69)$$

The method that we propose for estimating \mathbf{a} and \mathbf{b} is based on (68) and (69) and consists of two steps.

- Step 1) Estimate $\{\alpha_k\}_{k=1}^K$ from (69) in an unstructured/non-parametric form.
- Step 2) Fit $\{\alpha_k(\mathbf{a}, \mathbf{b})\}_{k=1}^K$ in (68) to the amplitude estimates obtained in the previous step by taking into account the statistical variance of the latter.

The motivation of this two-step method is as follows. The part of the negative log-likelihood function of the data in (69) that depends on the parameters of interest \mathbf{a} and \mathbf{b} is (under the Gaussian hypothesis) given by

$$[\mathbf{x} - \tilde{\mathbf{A}}\boldsymbol{\alpha}(\mathbf{a}, \mathbf{b})]^H \mathbf{W}^{-1} [\mathbf{x} - \tilde{\mathbf{A}}\boldsymbol{\alpha}(\mathbf{a}, \mathbf{b})] \quad (70)$$

where \mathbf{x} , $\tilde{\mathbf{A}}$, $\boldsymbol{\alpha}$, and \mathbf{W} are as defined in Section II-A. A simple calculation shows that we can rewrite (70) in the equivalent form

$$[\tilde{\boldsymbol{\alpha}} - \boldsymbol{\alpha}(\mathbf{a}, \mathbf{b})]^H (\tilde{\mathbf{A}}^H \mathbf{W}^{-1} \tilde{\mathbf{A}}) [\tilde{\boldsymbol{\alpha}} - \boldsymbol{\alpha}(\mathbf{a}, \mathbf{b})] + \text{constant}. \quad (71)$$

Here, $\tilde{\boldsymbol{\alpha}}$ is the maximum-likelihood estimate of the (unstructured) amplitude vector when the noise covariance matrix \mathbf{W} is given

$$\tilde{\boldsymbol{\alpha}} = (\tilde{\mathbf{A}}^H \mathbf{W}^{-1} \tilde{\mathbf{A}})^{-1} (\tilde{\mathbf{A}}^H \mathbf{W}^{-1} \mathbf{x}). \quad (72)$$

Next, we note that any of the amplitude estimators previously described provides a feasible large-sample realization of $\tilde{\boldsymbol{\alpha}}$ above, which does not require the knowledge of \mathbf{W} . Let $\hat{\boldsymbol{\alpha}}$ denote an estimate obtained by such an estimator. Since the estimation errors in $\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}$ vanish, as N increases, at a higher rate than the errors in $\tilde{\boldsymbol{\alpha}} - \boldsymbol{\alpha}$, we can replace $\tilde{\boldsymbol{\alpha}}$ in (71) with $\hat{\boldsymbol{\alpha}}$ without affecting the asymptotic properties of the estimates of \mathbf{a} and \mathbf{b} obtained by minimizing (71). For sufficiently large N , we can also replace the weighting matrix $\tilde{\mathbf{A}}^H \mathbf{W}^{-1} \tilde{\mathbf{A}}$ in (71) by a consistent estimate of its large-sample limit in (10), once again without affecting the asymptotics. By doing so, we obtain the following large-sample approximation of the criterion in (71):

$$C_1(\mathbf{a}, \mathbf{b}) = \sum_{k=1}^K \frac{1}{\hat{\phi}(\omega_k)} |\hat{\alpha}_k - \alpha_k(\mathbf{a}, \mathbf{b})|^2 \quad (73)$$

where $\hat{\phi}(\omega_k)$ is a consistent estimate of $\phi(\omega_k)$. To derive estimates of \mathbf{a} and \mathbf{b} as outlined above, we first need to compute $\hat{\boldsymbol{\alpha}}$ [Step 1] of the two-step method] and next minimize the fitting criterion in (73) with respect to \mathbf{a} and \mathbf{b} [Step 2)]. The so-obtained estimates of the parameters of interest in \mathbf{a} and \mathbf{b} are *asymptotically statistically efficient*, as implied by the above

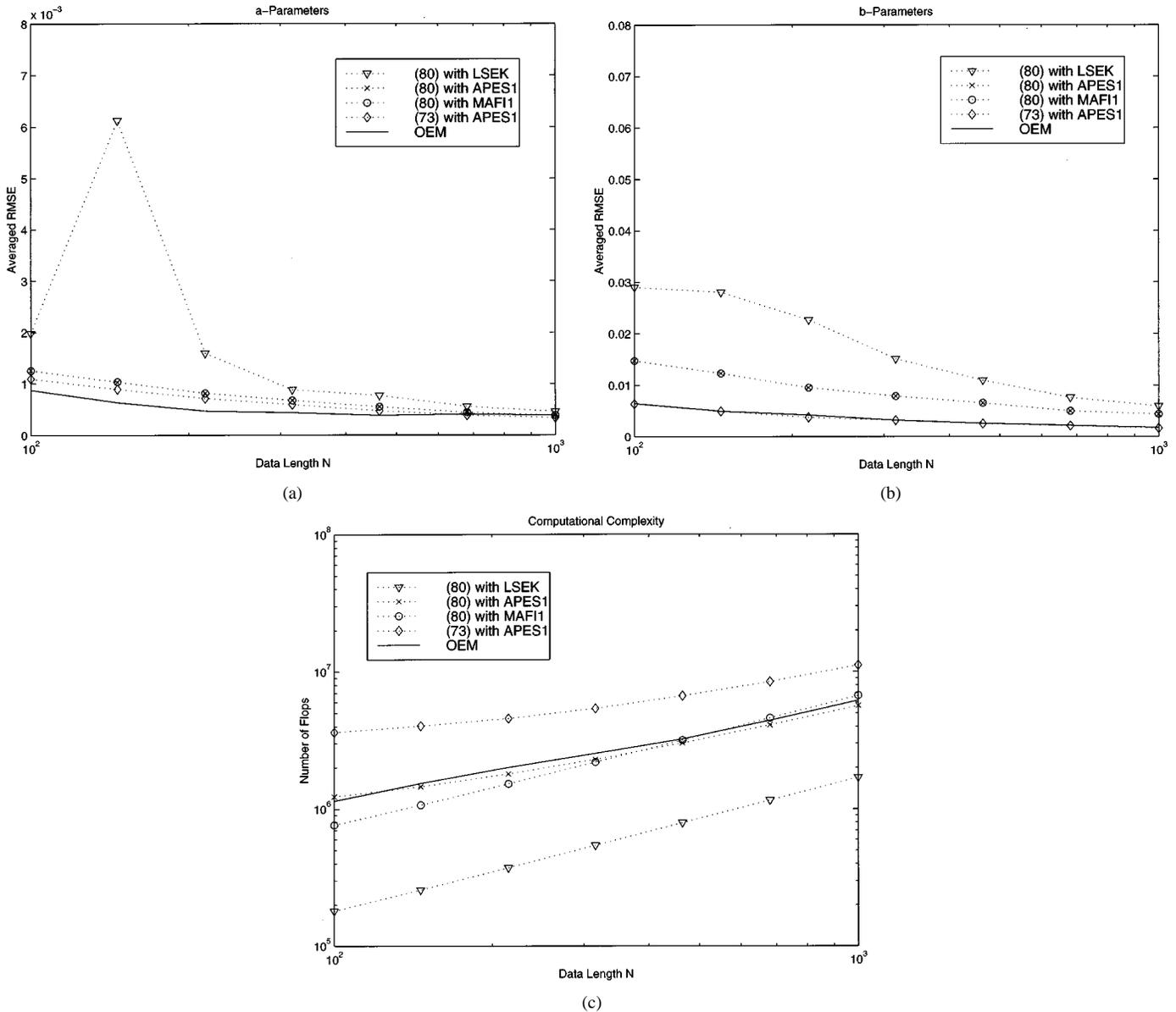


Fig. 7. Averaged RMSE's and the number of flops versus N for the first system when the observation noise is white ($\sigma^2 = 0.01$) and $M = 20$ for APES1 and MAF11. (a) RMSE of a parameters. (b) RMSE of b parameters. (c) Number of flops.

sketched derivation of (73). This fact also follows, in more exact terms, from the extended invariance principle (EXIP) [13].

As mentioned above, the estimates of \mathbf{a} and \mathbf{b} derived from (73) *asymptotically* achieve the CRB performance regardless of which of the amplitude estimators in Sections III and IV is used to obtain $\{\hat{\alpha}_k\}$. However, enhanced *finite-sample* performance will result if the amplitude estimator is carefully selected, for instance, by using the guidelines provided in the previous sections.

In what follows, we detail the two steps of the proposed method.

Step 1) Use an appropriate amplitude estimator to obtain estimates $\{\hat{\alpha}_k\}_{k=1}^K$ of $\{\alpha_k\}_{k=1}^K$ from the measurements $\{x(n)\}_{n=0}^{N-1}$. APES1 may be recommended in this case because we have control over the probing signal, and usually, we have no reason to choose any of the sinusoids too close to one another. The

large-sample variance of the estimated amplitudes $\{\hat{\alpha}_k\}_{k=1}^K$ is proportional to $\{\phi(\omega_k)\}_{k=1}^K$ (see Section II). To obtain the estimates of $\{\phi(\omega_k)\}_{k=1}^K$ needed in (73), we can first calculate

$$\hat{v}(n) = x(n) - \sum_{k=1}^K \hat{\alpha}_k e^{j\omega_k n} \quad n=0, 1, \dots, N-1 \quad (74)$$

and then utilize either a parametric or a non-parametric PSD estimator [5], [7], [8] to obtain $\{\hat{\phi}(\omega_k)\}_{k=1}^K$. In the examples given in Section VI-B, we use the Capon PSD estimator [2], [3], [5], which determines $\{\hat{\phi}(\omega_k)\}_{k=1}^K$ as

$$\hat{\phi}(\omega_k) = \frac{M}{\mathbf{a}^H(\omega_k) \hat{\mathbf{R}}_{\hat{v}}^{-1} \mathbf{a}(\omega_k)} \quad k=1, 2, \dots, K \quad (75)$$

where $\mathbf{a}(\omega_k)$ is defined in (34), and $\hat{\mathbf{R}}_{\hat{\mathbf{v}}\hat{\mathbf{v}}}$ is the sample covariance matrix of the estimated noise vectors

$$\hat{\mathbf{v}}(l) = [\hat{v}(l) \quad \hat{v}(l+1) \quad \cdots \quad \hat{v}(l+M-1)]^T$$

$$l=0, 1, \dots, L-1 \quad (76)$$

that is

$$\hat{\mathbf{R}}_{\hat{\mathbf{v}}\hat{\mathbf{v}}} = \frac{1}{L} \sum_{l=0}^{L-1} \hat{\mathbf{v}}(l) \hat{\mathbf{v}}^H(l). \quad (77)$$

Step 2) Obtain estimates of $\{a_i, b_j\}$ by minimizing the criterion in (73). To do so, we can use a host of methods, provided that we have good initial estimates of \mathbf{a} and \mathbf{b} .

To obtain initial estimates and then minimize (73), we assume that p and q are known. (Standard techniques for system order determination can be found in, e.g., [1] and [14].) We pick up the $p+q$ largest $\{\hat{\alpha}_k\}$ and define a criterion made from the corresponding terms of (73)

$$C_2(\mathbf{a}, \mathbf{b}) = \sum_{k=1}^{p+q} \frac{1}{\hat{\phi}(\omega_k)} |\hat{\alpha}_k - \alpha_k(\mathbf{a}, \mathbf{b})|^2 \quad (78)$$

where we have assumed, for notational simplicity, that $\{\hat{\alpha}_k\}_{k=1}^{p+q}$ are the $p+q$ chosen amplitudes. [Given that $\hat{\phi}(\omega_k)$ was estimated, an alternative would be to choose those $\{\hat{\alpha}_k\}$ that have the largest ratio $|\hat{\alpha}_k|^2 / \hat{\phi}(\omega_k)$; however, for the sake of computational simplicity, we may want to skip the estimation of $\phi(\omega_k)$; then, this alternative cannot be used. See below.] Now, the minimization of (78) is simple. Indeed, we can choose \mathbf{a} and \mathbf{b} to satisfy

$$\hat{\alpha}_k = \alpha_k(\mathbf{a}, \mathbf{b}), \quad k = 1, 2, \dots, p+q \quad (79)$$

or, equivalently

$$\frac{\hat{\alpha}_k}{\gamma_k} A(e^{-j\omega_k}) = B(e^{-j\omega_k}), \quad k = 1, 2, \dots, p+q \quad (80)$$

which can be rewritten as a linear system of $p+q$ equations with $p+q$ unknowns $\{a_i, b_j\}$. That system will generally have a unique solution that is also the minimizer of (78) [it makes (78) equal to zero], and which gives our initial estimates of $\{a_i, b_j\}$. As shown in the following numerical examples, *the initial estimates are usually quite good. Hence, we can skip the step of minimizing (73) to save computations.*

We reiterate that according to the aforementioned EXIP of [13], *the estimates of $\{a_i, b_j\}$ obtained by minimizing (73) achieve the CRB asymptotically, and hence, they have a better asymptotic accuracy than the OEM estimates whenever $v(n)$ is colored.* It also follows from this observation that in the case of $K = p+q$, the simple estimates obtained from (80) are asymptotically efficient. This latter result (of a somewhat limited interest due to the requirement that $K = p+q$) was first proved in [15] in a relatively complicated way.

B. Numerical Examples

The following examples assume that p and q are known to facilitate performance comparison. It is reasonable to do so since both OEM and the proposed method use similar techniques to

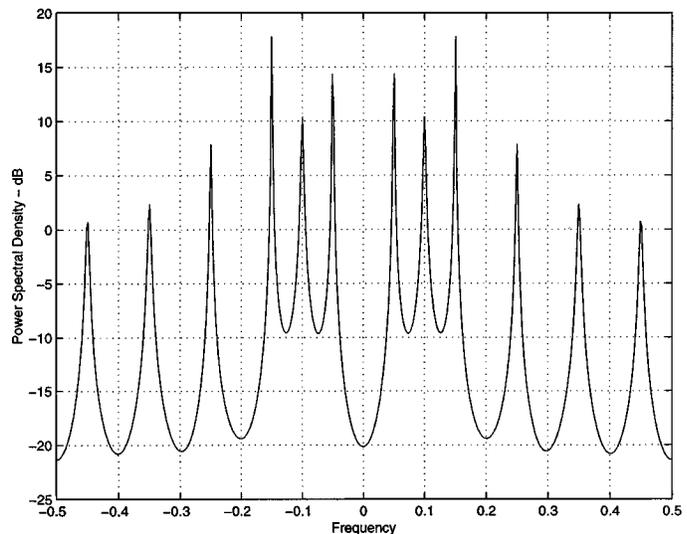


Fig. 8. PSD estimate of the output of the first system corrupted by white noise with $\sigma^2 = 0.01$ and $N = 200$.

determine the model orders. In addition, we adopt the strategy to choose the $p+q$ largest $\{\hat{\alpha}_k\}$ to obtain the initial estimates of \mathbf{a} and \mathbf{b} needed in Step 2) of the proposed method.

Example 6.2.1: The system considered in this example is given by (65) with

$$A(z^{-1}) = 1 - 1.6019z^{-1} + 0.9801z^{-2} \quad (81)$$

and

$$B(z^{-1}) = z^{-1} + 0.2472z^{-2} + 0.1600z^{-3}. \quad (82)$$

The probing signal is given by

$$u(n) = 2 \cos(2\pi 0.05n) + 2 \cos(2\pi 0.15n) + 2 \cos(2\pi 0.25n)$$

$$+ 2 \cos(2\pi 0.35n) + 2 \cos(2\pi 0.45n)$$

$$n = 0, 1, \dots, N-1. \quad (83)$$

We consider using a real-valued probing signal because this is the usual case in practice. (A subtle question then arises as the amplitude estimation techniques discussed in the previous sections all assume that the sinusoids are complex valued. We might impose certain conjugate symmetry constraints and derive similar techniques that are specifically tailored to real-valued sinusoidal amplitude estimation so that if, for example $\omega_1 = -\omega_2$, then the estimators will give $\hat{\alpha}_1 = \hat{\alpha}_2^*$. Yet, our experience shows that the gain would most often be minor, and hence, the effort is not worthwhile. See [16], for example.) Note that $K = 10$ for the signal in (83). The noise $v(n)$ is a real-valued *white* Gaussian noise with zero-mean and variance $\sigma^2 = 0.01$. We estimate the system parameters using the proposed technique and *the (time-domain) OEM* that obtains parameter estimates by minimizing (67). (The computer routine used for OEM is the one provided in the system identification toolbox of MATLAB). For the proposed technique, we compute both *the initial estimates given by solving (80) and the minimizer of (73)*. The minimizer of (73) is found by using the solution of (80), based on APES1, as the initial condition and then employing a

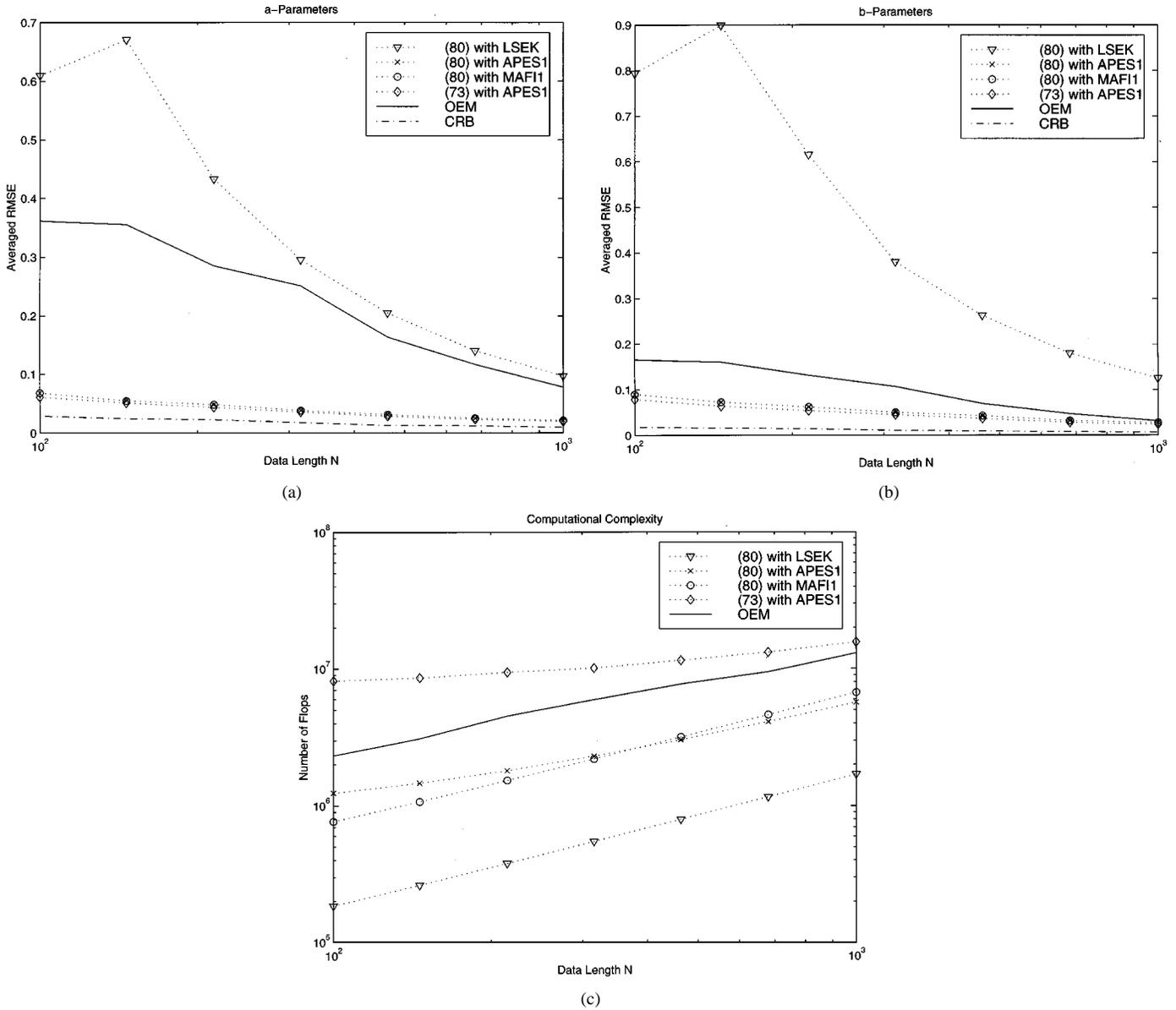


Fig. 9. Averaged RMSE's and the number of flops versus N for the second system when the observation noise is colored [an AR(1) process with $\sigma^2 = 0.01$] and $M = 20$ for APES1 and MAFI1. (a) RMSE of a parameters. (b) RMSE of b parameters. (c) Number of flops.

standard gradient-type nonlinear optimization routine provided by MATLAB. To reduce the number of graphs, we only show the averaged root mean squared error (RMSE) for the a parameters

$$\text{RMSE}\{\hat{\mathbf{a}}\} = \frac{1}{p} \sum_{i=1}^p \text{RMSE}\{\hat{a}_i\} \quad (84)$$

and similarly for the b parameters. All results are based on 200 Monte Carlo simulations. Fig. 7(a) and (b) show the averaged RMSE's of the a parameters and, respectively, the b parameters obtained by using OEM and the proposed technique, as N increases. Fig. 7(c) shows the required number of flops as N increases. (APES1 and MAFI1 use $M = 20$ in this and the following example, which does not fall in the range given in Table I. The reason is that APES1 or MAFI1 with $M = 20$ is quite acceptable for the probing signal in (83) that contains well-sep-

arated sinusoids, and hence, choosing a larger M would only result in additional computations.) Finding the minimizer of (73) or the OEM estimates involve iterative searches that give variable flop counts from trial to trial. The number of flops needed by each of these two methods, as shown in Fig. 7(c), is the average over 200 trials. As we can see, the initial estimates of $\{a_i, b_j\}$ given by using (80) with APES1 or MAFI1 have similar RMSE's to those obtained by OEM. The estimates obtained by minimizing (73) are slightly better than the initial estimates obtained using APES1 or MAFI1 but at a significantly increased computational cost. Due to this observation, *we do not recommend using this approach, i.e., minimizing (73), for refined estimation accuracy.* Fig. 7(c) also shows that as compared with OEM, there is little computational advantage associated with using the initial estimates obtained by APES1 and MAFI1. The reason may be that the system in this example is quite simple (it has white output errors, etc.), and apparently, OEM reaches con-

vergence in a relatively small number of iterations. For a more complex system, such as the one used in the next example, OEM may need more iterations to converge. It should also be mentioned that we did not program our method very carefully, and hence, our code is unlikely to be as efficient as the OEM code in MATLAB. Regarding the estimation accuracy, we will stress that in the current case where the noise $v(n)$ is white, OEM coincides with the optimal maximum likelihood method (MLM) [1], [14]. When $v(n)$ is colored, OEM is no longer MLM. In that case, the initial system parameter estimates obtained by APES1 or MAFI1 may outperform OEM, as shown in the next example.

Recall that LSEK is statistically efficient when the observation noise is white. Then, we might wonder why the initial estimates given by LSEK may be notably worse in such a case than those given by APES1 or MAFI1, as happened in the previous example (especially when N is small). The reason is that the transient response of this system cannot be neglected for small N . To show this, the PSD of $x(n)$ is estimated by using the Capon PSD estimator, with $N = 200$ and $M = 20$, and is plotted in Fig. 8. It shows two extra peaks (which behave like two sinusoids) at ± 0.1 . The extra peaks are attributed to the response of the system (which has poles at $0.99e^{\pm j2\pi 0.1}$) to the initial conditions. Since it is essential for the LSEK to have *accurate* knowledge of the number and frequencies of the sinusoids to give reliable amplitude estimates, there is no surprise (with the above explanation in mind) that its performance in the previous example is considerably deteriorated. On the other hand, the above knowledge is not necessarily needed by APES1, and hence, its performance is not affected. Unlike APES1, MAFI1 does require this knowledge. Yet, the system response to the initial condition is substantially weakened by the frequency selective filtering employed by MAFI1, and hence, it has little, if any, effect on the amplitude estimates and the system parameter estimates. As N increases, the transient effect becomes less severe, and consequently, the initial system parameter estimates obtained by using LSEK approach those obtained by APES1 or MAFI1.

Example 6.2.2: We now consider a second system with

$$A(z^{-1}) = 1 - 1.9109z^{-1} + 1.7251z^{-2} - 0.7033z^{-3} + 0.2450z^{-4} \quad (85)$$

and

$$B(z^{-1}) = z^{-1} + 1.0562z^{-2} + 0.6100z^{-3} + 0.1912z^{-4} + 0.0400z^{-5}. \quad (86)$$

The noise $v(n)$ is an *AR(1)* signal as in (61), except that $e(n)$ is now replaced by a real-valued white Gaussian noise with zero-mean and variance $\sigma^2 = 0.01$. The probing signal is the same as in the previous example. Fig. 9(a)–(c) show the averaged RMSE's of the a parameters and the b parameters, as well as the number of flops, as N increases. As we can see, the initial system parameter estimates given by APES1 or MAFI1 are nearly statistically efficient and significantly better than those given by OEM, and yet, the former two are computationally more efficient than the latter. *Our conclusion is that the method*

that derives estimates of \mathbf{a} and \mathbf{b} by solving the linear system in (80), where $\{\hat{\alpha}_k\}$ are obtained by either APES1 or MAFI1, should be the method of choice for the system identification problem under discussion.

VII. CONCLUSION

In this paper, we investigated the problem of amplitude estimation of sinusoidal signals in colored noise. Three general classes of estimators, namely the LS, WLS, and MAFI approaches to amplitude estimation, have been discussed. We have shown that under certain circumstances, the MAFI approach to amplitude estimation is equivalent to the WLS approach, and yet, the former is more general and includes the latter as a special case. The amplitude estimators under discussion can be further categorized, depending on whether they estimate one amplitude at a time or all amplitudes simultaneously. MAFI or WLS methods, such as APES1, APESK, and MAFI1, in general give more accurate amplitude estimates for sinusoids in colored noise. Methods that estimate only one amplitude at a time, such as APES1, do not necessarily require the exact knowledge of the number and locations of the sinusoids and, hence, are more robust than those that estimate all amplitudes simultaneously.

We have also studied a system identification application using sinusoidal probing signals. We have presented a simple technique for system identification that can avoid iterative search such as the one required by OEM. We have shown that by using this simple technique with appropriate amplitude estimators, such as APES1 or MAFI1, we can obtain results that are generally better than those corresponding to the widely used iterative OEM, yet usually at a reduced computational cost.

REFERENCES

- [1] T. Söderström and P. Stoica, *System Identification*, London, U.K.: Prentice-Hall, 1989.
- [2] J. Capon, "High resolution frequency-wavenumber spectrum analysis," *Proc. IEEE*, vol. 57, pp. 1408–1418, Aug. 1969.
- [3] R. T. Lacoss, "Data adaptive spectral analysis methods," *Geophysics*, vol. 36, no. 4, pp. 661–675, Aug. 1971.
- [4] J. Li and P. Stoica, "An adaptive filtering approach to spectral estimation and SAR imaging," *IEEE Trans. Signal Processing*, vol. 44, pp. 1469–1484, June 1996.
- [5] P. Stoica and R. L. Moses, *Introduction to Spectral Analysis*. Upper Saddle River, NJ: Prentice-Hall, 1997.
- [6] P. Stoica, A. Jakobsson, and J. Li, "Matched-filter bank interpretation of some spectral estimators," *Signal Process.*, vol. 66, no. 1, pp. 45–59, Apr. 1998.
- [7] S. M. Kay, *Modern Spectral Estimation: Theory and Application*. Englewood Cliffs, NJ: Prentice-Hall, 1988.
- [8] S. L. Marple, Jr., *Digital Spectral Analysis with Applications*. Englewood Cliffs, NJ: Prentice-Hall, 1987.
- [9] U. Grenander and M. Rosenblatt, *Statistical Analysis of Stationary Time Series*, Stockholm, Sweden: Almqvist och Wiksell, 1956.
- [10] E. J. Hannan and B. Wahlberg, "Convergence rates for inverse Toeplitz matrix forms," *J. Multivariate Anal.*, vol. 31, pp. 127–135, Oct. 1989.
- [11] H. Li, J. Li, and P. Stoica, "Performance analysis of forward-backward matched-filterbank spectral estimators," *IEEE Trans. Signal Processing*, vol. 46, pp. 1954–1966, July 1998.
- [12] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1985.
- [13] P. Stoica and T. Söderström, "On reparameterization of loss functions used in estimation and the invariance principle," *Signal Process.*, vol. 17, pp. 383–387, Aug. 1989.
- [14] L. Ljung, *System Identification: Theory for the User*. Englewood Cliffs, NJ: Prentice-Hall, 1987.

- [15] P. V. Kabaila, "On output-error methods for system identification," *IEEE Trans. Automat. Contr.*, vol. AC-28, pp. 12–23, Jan. 1983.
- [16] A. Jakobsson, T. Ekman, and P. Stoica, "Capon and APES spectrum estimation for real-valued signals," presented at the 8th IEEE DSP Workshop, Bryce Canyon, UT, Aug. 1998.



Petre Stoica (F'94) received the D.Sc. degree from the Bucharest Polytechnic Institute, Bucharest, Romania, in 1979, and an Honorary Doctorate degree from Uppsala University, Uppsala, Sweden, in 1993.

He is a Professor of System Modeling with the Department of Systems and Control at Uppsala University. Previously, he was a Professor of Signal Processing at the Bucharest Polytechnic Institute. He held longer visiting positions with the Eindhoven University of Technology, Eindhoven, The Netherlands, Chalmers Institute of

Technology, Gothenburg, Sweden, Uppsala University, the University of Florida, Gainesville, and Stanford University, Stanford, CA. His main scientific interests are in the areas of system identification, time series analysis and prediction, statistical signal and array processing, spectral analysis, wireless communications, and radar signal processing. He has published seven books, eight book chapters, and some 370 papers in archival journals and conference records on these topics. His most recent book is *Introduction to Spectral Analysis*, with R. Moses (Englewood Cliffs, NJ: Prentice-Hall, 1997).

Dr. Stoica was co-recipient of the 1989 ASSP Senior Award and recipient of the 1996 Technical Achievement Award of the IEEE Signal Processing Society. He is on the Editorial Boards of five journals in the field: *Signal Processing*; *Journal of Forecasting*; *Circuits, Systems and Signal Processing*; *Multidimensional Systems and Signal Processing*; and *Digital Signal Processing: A Review Journal*. He was Guest Co-editor for several special issues on system identification, spectral analysis, and radar. He is an Honorary Member of the Romanian Academy and a Fellow of the Royal Statistical Society.



Hongbin Li (S'98–M'99) received the B.S. and M.S. degrees from the University of Electronic Science and Technology of China, Chengdu, in 1991 and 1994, respectively, and the Ph.D. degree from the University of Florida, Gainesville, in 1999, all in electrical engineering.

From July 1996 to May 1999, he was a Research Assistant with the Department of Electrical and Computer Engineering, University of Florida. Since July 1999, he has been an Assistant Professor with the Department of Electrical and Computer Engineering,

Stevens Institute of Technology, Hoboken, NJ. His current research interests include stochastic signal processing, sensor array processing, wireless communications, and radar imaging.

Dr. Li is a member of Tau Beta Pi and Phi Kappa Phi. He received the 1999 Sigma Xi Graduate Research Award from the University of Florida Chapter.



Jian Li (S'87–M'91–SM'97) received the M.Sc. and Ph.D. degrees in electrical engineering from The Ohio State University (OSU), Columbus, in 1987 and 1991, respectively.

From April 1991 to June 1991, she was an Adjunct Assistant Professor with the Department of Electrical Engineering, OSU. From July 1991 to June 1993, she was an Assistant Professor with the Department of Electrical Engineering, University of Kentucky, Lexington. Since August 1993, she has been with the Department of Electrical and Computer Engineering, University of Florida, Gainesville, where she is currently an Associate Professor. Her current research interests include spectral estimation, synthetic aperture radar image formation and understanding, radar detection and estimation theory, sensor array signal processing, image segmentation and processing, and communications.

Dr. Li is a member of Sigma Xi and Phi Kappa Phi. She received the 1994 National Science Foundation Young Investigator Award and the 1996 Office of Naval Research Young Investigator Award.