2-D Sinusoidal Amplitude Estimation with Application to 2-D System Identification *

Hongbin Li† Wei Sun† Petre Stoica‡ Jian Li§

December 14, 2001

Point of Contact: Dr. Hongbin Li, Department of Electrical and Computer Engineering, Stevens Institute of Technology, Castle Point on Hudson, Hoboken, NJ 07030, USA. Phone: (201) 216 5604. Fax: (201) 216 8246. e-mail: hli@stevens-tech.edu.

Abstract

In a companion paper [1], we studied amplitude estimation of one-dimensional (1-D) sinusoidal signals from measurements corrupted by possibly colored observation noise. We herein extend the results for two-dimensional (2-D) amplitude estimation, which is of interest in various applications, including medical imaging, synthetic aperture radar (SAR), seismology, and many others. In particular, we investigate 2-D sinusoidal amplitude estimation under the general frameworks of least-squares (LS), weighted-least-squares (WLS), and MAtched-FIlterbank (MAFI) estimation. A variety of 2-D amplitude estimators are presented, which do not model the observation noise exactly but are all asymptotically (for large samples) statistically efficient. The performances of these estimators in finite samples are compared numerically with one another as well as with the Cramér-Rao bound (CRB), the lower variance bound for any unbiased estimators. Making use of amplitude estimation techniques, we introduce a new scheme for 2-D system identification, which has a closed-form expression. The proposed 2-D system identification scheme is shown to be computationally simpler and statistically more accurate than the conventional output error method (OEM), when the observation noise is colored. The CRB for the 2-D system identification problem is also investigated in this paper. Close-to-CRB performances are observed for the proposed system identification scheme for both white and colored noise with moderate numbers of data samples.

*This work was supported by the New Jersey Center for Wireless Telecommunications (NJCWT), the Senior Individual Grant Program of the Swedish Foundation for Strategic Research, and the National Science Foundation under Grant MIP-9457388.

†Hongbin Li and Wei Sun are with the Department of Electrical and Computer Engineering, Stevens Institute of Technology, Hoboken, NJ 07030, USA (e-mail: {hli,wsun}@stevens-tech.edu).

‡Petre Stoica is with the Department of Systems and Control, Uppsala University, P.O.Box 27, SE–751 03, Uppsala, Sweden (e-mail: ps@syscon.uu.se).

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

Key Words

two-dimensional (2-D) amplitude estimation, 2-D spectral analysis, 2-D system identification, least squares, weighted least squares, Cramér-Rao bound.

1 Introduction

The need to estimate the complex amplitudes of sinusoidal signals in a noisy environment is encountered in many signal processing applications [2], [3]. In radars, for example, the complex amplitude of the demodulated signal, also referred to as the target range signature, is often used for radar target detection and recognition (see, e.g., [4] and references therein).

A null filtering technique making use of a statistically optimal null filter (SONF) was recently proposed in [5] for amplitude estimation of arbitrary waveforms. The SONF combines instantaneous matched filtering with least-squares (LS) estimation, and can be used for *joint* and *separate* estimation of the amplitudes of multiple waveforms. It turns out that the joint and separate amplitude estimators via SONF for sinusoidal signals coincide with two LS-based amplitude estimators, referred to as LSE$(1, 0, 1)$ and LSE$(1, 0, K)$, respectively, in [1] [also see Sections 3.1.1 and 3.1.2 of this paper for the two-dimensional (2-D) versions]. When the observation noise is spectrally *white* and the signal-to-noise ratio (SNR) is *known*, minimum-mean-squared-error (MMSE) amplitude estimation via the SONF is possible; the MMSE estimates, however, are biased.

LS-based amplitude estimators, in effect, have been most widely used due to their conceptual and computational simplicity. LS amplitude estimates are also known to be optimum when the observation noise follows a white Gaussian distribution. Despite their popularity, LS amplitude estimates are far from being satisfactory when the observation noise is *colored* and, particularly, when the size of the observed data is relatively *small* [1]. In view of this, we investigated alternative techniques for amplitude estimation, including weighted-least-squares (WLS) and MAtched-FIlterbank (MAFI) approaches in [1]. It was shown that WLS and MAFI based techniques lead to statistically more accurate amplitude estimates than the LS estimates when the observation noise is colored; meanwhile, when the observation noise *is* white, in which case the LS approach is optimal, WLS and MAFI based amplitude estimators still attain close-to-optimal performances. All such properties make WLS and MAFI amplitude estimators very desirable, especially when the observation noise is colored, or when no *a priori* information about the spectrum of the observation noise is available.

Our study in [1], however, was limited to one-dimensional (1-D) signals. As many applications, includ-

ing synthetic aperture radar (SAR) imaging, medical imaging, seismic signal processing, etc., are concerned with 2-D amplitude estimation, it would be of substantial interest to extend the results in [1] to 2-D scenarios. In this paper, we explore such an extension. As will be seen in the sequel, non-trivial efforts are required to extend the WLS and MAFI techniques to amplitude estimation of 2-D sinusoidal signals. Aside from 2-D amplitude estimation, we also investigate a 2-D system identification problem. By utilizing 2-D amplitude estimation techniques, we propose a new approach leading to a simple closed-form solution to the 2-D system identification problem.

The rest of the paper is organized as follows. In Section 2, we formulate the problem of interest. In Section 3, we investigate various ways for 2-D amplitude estimation. Specifically, in Section 3.1, we discuss the LS approaches to 2-D amplitude estimation. That section also contains a statistical analysis of LS based amplitude estimators. In Section 3.2, we introduce several WLS based 2-D amplitude estimators. These amplitude estimators are then linked to the classical Capon [6] and recently proposed APES (Amplitude and Phase EStimation) [4], [7] spectral estimators, both obtaining spectral estimates from adaptively filtered outputs of the observed data. The connection between parameter estimation and adaptive filtering brings up a more general class of MAFI based amplitude estimators, a topic discussed in Section 3.3. The MAFI concept was originally proposed in [8] and [9] for spectral analysis. Here, we extend the idea to 2-D amplitude estimation. Numerical examples which compare the performances of all these 2-D amplitude estimators are presented in Section 4. As an application example, we describe in Section 5 how to use 2-D amplitude estimation techniques to solve the 2-D system identification problem. The proposed solution has a closed-form and is computationally simpler and statistically more accurate than the conventional output error method (OEM) when the observation is colored. Finally, the paper is concluded in Section 6.

2 Problem Formulation

Consider the noise-corrupted observation of K 2-D complex-valued sinusoids

$$x(n,\bar{n}) = \sum_{k=1}^{K} \alpha_k e^{j2\pi(f_k n + \bar{f}_k \bar{n})} + v(n,\bar{n}), \quad n = 0,\cdots,N-1;\ \bar{n} = 0,\cdots,\bar{N}-1, \tag{1}$$

where $\{\alpha_k\}_{k=1}^K$ is the complex amplitude of the k-th 2-D complex sinusoid at the 2-D frequency pair $(f_k,\bar{f}_k)$, and $v(n,\bar{n})$ is the 2-D complex-valued additive observation noise assumed to be stationary with zero-mean and unknown finite power spectral density (PSD) $\phi(f,\bar{f})$. We assume that the frequency pairs, $\{f_k,\bar{f}_k\}_{k=1}^K$, are known and distinct from one another. The problem of interest is to estimate the 2-D amplitudes $\{\alpha_k\}_{k=1}^K$ from the observations $\{x(n,\bar{n})\}$.

In order to write (1) more compactly, we need a few definitions. Let

$$\mathbf{X} \stackrel{\triangle}{=} \begin{bmatrix} x(0,0) & \cdots & x(0,\bar{N}-1) \\ x(1,0) & \cdots & x(1,\bar{N}-1) \\ \vdots & \vdots & \vdots \\ x(N-1,0) & \cdots & x(N-1,\bar{N}-1) \end{bmatrix}, \tag{2}$$

and let $\mathbf{V}$ be built from $\{v(n,\bar{n})\}$ in the same manner as $\mathbf{X}$ is from $\{x(n,\bar{n})\}$. Using matrix notations, (1) can be expressed as

$$\mathbf{X} = \mathbf{A}\mathbf{\Lambda}\bar{\mathbf{A}}^T + \mathbf{V}, \tag{3}$$

where $(\cdot)^T$ denotes the matrix transpose, and

$$\mathbf{A} \stackrel{\triangle}{=} \begin{bmatrix} \boldsymbol{a}(f_1) & \cdots & \boldsymbol{a}(f_K) \end{bmatrix}, \quad \bar{\mathbf{A}} \stackrel{\triangle}{=} \begin{bmatrix} \bar{\boldsymbol{a}}(\bar{f}_1) & \cdots & \bar{\boldsymbol{a}}(\bar{f}_K) \end{bmatrix}, \quad \mathbf{\Lambda} \stackrel{\triangle}{=} \text{diag}\begin{Bmatrix} \alpha_1 & \cdots & \alpha_K \end{Bmatrix},$$

with the k-th column of $\mathbf{A}$ and, respectively, of $\bar{\mathbf{A}}$ given by

$$\boldsymbol{a}(f_k) = \begin{bmatrix} 1 & \cdots & e^{j(N-1)2\pi f_k} \end{bmatrix}^T, \quad \bar{\boldsymbol{a}}(\bar{f}_k) = \begin{bmatrix} 1 & \cdots & e^{j(\bar{N}-1)2\pi \bar{f}_k} \end{bmatrix}^T. \tag{4}$$

Alternatively, we can work with vectors instead of matrices. Let $\boldsymbol{x} \stackrel{\triangle}{=} \text{vec}\{\mathbf{X}\}$ and $\boldsymbol{v} \stackrel{\triangle}{=} \text{vec}\{\mathbf{V}\}$, where vec$\{\cdot\}$ stacks the columns of the matrix argument over one another [10]. We vectorize both sides of (3), which yields

$$\boldsymbol{x} = (\bar{\mathbf{A}} \diamond \mathbf{A})\boldsymbol{\alpha} + \boldsymbol{v} \stackrel{\triangle}{=} \mathbf{\Psi}\boldsymbol{\alpha} + \boldsymbol{v}, \tag{5}$$

where $\boldsymbol{\alpha} \stackrel{\triangle}{=} \begin{bmatrix} \alpha_1 & \cdots & \alpha_K \end{bmatrix}^T$ and $\diamond$ denotes the matrix Khatri-Rao product [11], a column-wise Kronecker product. That is, $\mathbf{\Psi}$ can be expressed as

$$\mathbf{\Psi} = \begin{bmatrix} \bar{\boldsymbol{a}}(\bar{f}_1) \otimes \boldsymbol{a}(f_1) & \ldots & \bar{\boldsymbol{a}}(\bar{f}_K) \otimes \boldsymbol{a}(f_K) \end{bmatrix}, \tag{6}$$

where $\otimes$ denotes the Kronecker product [10].

We will henceforth mainly use (5) instead of (3) for estimation and analysis. Accordingly, the problem of interest is equivalent to determining the amplitude vector $\boldsymbol{\alpha}$ from the data vector $\boldsymbol{x}$ when the PSD, $\phi(f,\bar{f})$, of $\boldsymbol{v}$ is unknown.

3 2-D Amplitude Estimators

In this section, we discuss a variety of 2-D amplitude estimators which can be categorized as least-squares (LS), weighted-least-squares (WLS), and MAtched-FIlterbank (MAFI) based 2-D amplitude estimators.

They can also be classified depending on whether they consider one sinusoid at a time or all sinusoids simultaneously, or whether pre-filtering is employed for amplitude estimation. Furthermore, some estimators split the data matrix into submatrices to artificially create multiple "snapshots", whereas the others do not. In order to discriminate these different amplitude estimators, we use the following naming convention. For example, LSE$(1,0,1)$ denotes the LS estimator that uses a single data snapshot (hence it does not split the data matrix into overlapping submatrices), employs no pre-filtering, and estimates one amplitude at a time. Likewise, MAFI$(L\bar{L},K,K)$ denotes the MAFI estimator that splits the data into $L\bar{L}$ submatrices [cf. (15) and (16)], utilizes a bank of K pre-filters, and estimates K amplitudes simultaneously. The remaining amplitude estimators are similarly designated.

3.1 Least-Squares Amplitude Estimators

Among all the amplitude estimators to be discussed in this paper, the LS estimators are most straightforward. We consider two LS estimators, namely LSE$(1,0,K)$ and LSE$(1,0,1)$. As both estimators do not take the correlation of the noise into account, they produce unsatisfactory results when the observation noise is colored.

3.1.1 LSE$(1,0,K)$

From (5), the LS estimate of $\boldsymbol{\alpha}$ is

$$\hat{\boldsymbol{\alpha}} = (\boldsymbol{\Psi}^H\boldsymbol{\Psi})^{-1}\boldsymbol{\Psi}^H\boldsymbol{x}, \tag{7}$$

where $(\cdot)^H$ denotes the conjugate transpose. It is readily verified that $\hat{\boldsymbol{\alpha}}$ is *unbiased*, i.e., $E\{\hat{\boldsymbol{\alpha}}\} = \boldsymbol{\alpha}$, where $E\{\cdot\}$ denotes statistical expectation. The mean squared error (MSE) of $\hat{\boldsymbol{\alpha}}$ is given by

$$\text{MSE}\{\hat{\boldsymbol{\alpha}}\} = (\boldsymbol{\Psi}^H\boldsymbol{\Psi})^{-1}\boldsymbol{\Psi}^H\boldsymbol{\Gamma}\boldsymbol{\Psi}(\boldsymbol{\Psi}^H\boldsymbol{\Psi})^{-1}, \tag{8}$$

where $\boldsymbol{\Gamma} \triangleq E\{\boldsymbol{v}\boldsymbol{v}^H\}$ is the covariance matrix of the noise. Note that the matrix $\boldsymbol{\Gamma}$ is Hermitian and block Toeplitz.

Under the mild assumption that $v(n,\bar{n})$ is circularly symmetric Gaussian, the Cramér-Rao bound (CRB) for $\hat{\boldsymbol{\alpha}}$ can be straightforwardly obtained by using, for example, the Slepian-Bangs formula (see, e.g., [12]):

$$\text{CRB}(\boldsymbol{\alpha}) = (\boldsymbol{\Psi}^H\boldsymbol{\Gamma}^{-1}\boldsymbol{\Psi})^{-1}. \tag{9}$$

We can verify that MSE$\{\hat{\boldsymbol{\alpha}}\}$ = CRB$\{\boldsymbol{\alpha}\}$ when the observation noise is white, i.e., $\boldsymbol{\Gamma} \sim \mathbf{I}_{N\bar{N}}$, where $\mathbf{I}_{N\bar{N}}$ is the $N\bar{N} \times N\bar{N}$ identity matrix, *and* $N\bar{N} \geq K$. That is, *with white noise, LSE*$(1,0,K)$ *is statistically optimal*. In general, when the observation noise is colored, we have MSE$\{\hat{\boldsymbol{\alpha}}\} >$ CRB$\{\boldsymbol{\alpha}\}$, that is, LSE$(1,0,K)$ is statistically inefficient with colored noise.

Despite its aforementioned statistical inefficiency for finite data samples, LSE$(1,0,K)$ is asymptotically (for large samples) statistically efficient, as summarized in the following result.

Theorem 1 *Under the conditions that the observed data can be described as in (1), and that the observation noise $v(n,\bar{n})$ is stationary with zero-mean and finite PSD $\phi(f,\bar{f})$, LSE$(1,0,K)$ is asymptotically (for large N and $\bar{N}$) statistically efficient, and the asymptotic MSE of $\hat{\boldsymbol{\alpha}}$ is given by*

$$\lim_{N,\bar{N}\to\infty} N\bar{N}MSE\{\hat{\boldsymbol{\alpha}}\} = \lim_{N,\bar{N}\to\infty} N\bar{N}CRB\{\boldsymbol{\alpha}\} = \mathrm{diag}\left\{\phi(f_1,\bar{f}_1) \quad \cdots \quad \phi(f_K,\bar{f}_K)\right\}. \tag{10}$$

Proof: See Appendix A. □

3.1.2 LSE$(1,0,1)$

If we use LS to estimate one sinusoid at a time, and treat the rest $(K-1)$ sinusoids in (1) as noise, we obtain another LS amplitude estimator, namely LSE$(1,0,1)$:

$$\hat{\alpha}_k = \frac{1}{N\bar{N}} \sum_{n=0}^{N-1} \sum_{\bar{n}=0}^{\bar{N}-1} x(n,\bar{n}) e^{-j2\pi(f_k n + \bar{f}_k \bar{n})}, \quad k = 1, \cdots, K, \tag{11}$$

which is recognized as the normalized 2-D discrete Fourier transform (DFT) of $x(n,\bar{n})$ at frequency $(f_k,\bar{f}_k)$. Apparently, LSE$(1,0,1)$ is computationally much simpler than LSE$(1,0,K)$. Another advantage enjoyed by LSE$(1,0,1)$ (and by other one-at-a-time type of estimators to be discussed in the sequel) is that the *exact* knowledge of the 2-D frequencies of the sinusoids is *not* necessary. One can estimate both $\{\alpha_k\}_{k=1}^K$ and $\{f_k,\bar{f}_k\}_{k=1}^K$ by estimating one amplitude at a time for varying frequency over the frequency bands of interest and, then, identifying the peaks in the so-obtained spectrum [2], [3], [12]. This gives one-at-a-time estimators an extra flexibility when compared to their all-at-once counterparts. Note, however, that these advantages are obtained at the cost of degraded performance when some of the sinusoids are closely spaced to one another, as will be seen in Section 4.

An analysis of LSE$(1,0,1)$ proceeds as follows. Without loss of generality, we consider (11) for $k=1$. Let $\tilde{\boldsymbol{a}}_1 \stackrel{\triangle}{=} \bar{\boldsymbol{a}}(\bar{f}_1) \otimes \boldsymbol{a}(f_1)$. The LSE$(1,0,1)$ estimate of α_1 can be written as

$$\hat{\alpha}_1 = \left(\tilde{\boldsymbol{a}}_1^H \tilde{\boldsymbol{a}}_1\right)^{-1} \tilde{\boldsymbol{a}}_1^H \boldsymbol{x}. \tag{12}$$

Let $\boldsymbol{\beta} \stackrel{\triangle}{=} \begin{bmatrix} \alpha_2 & \dots & \alpha_K \end{bmatrix}^T$ and $\boldsymbol{\Xi}$ be obtained from $\boldsymbol{\Psi}$ by removing the first column of $\boldsymbol{\Psi}$ [cf. (6)], that is, $\boldsymbol{\Psi} = \begin{bmatrix} \tilde{\boldsymbol{a}}_1 & \boldsymbol{\Xi} \end{bmatrix}$. Taking the expectation of both sides of (12) yields

$$E\{\hat{\alpha}_1\} = \left(\tilde{\boldsymbol{a}}_1^H \tilde{\boldsymbol{a}}_1\right)^{-1} \tilde{\boldsymbol{a}}_1^H \boldsymbol{\Psi}\boldsymbol{\alpha} = \alpha_1 + \frac{1}{N\bar{N}} \tilde{\boldsymbol{a}}_1^H \boldsymbol{\Xi}\boldsymbol{\beta},$$

which shows that LSE$(1,0,1)$ is a biased estimator. Yet, it is asymptotically unbiased as $N \to \infty$ or $\bar{N} \to \infty$. The MSE of $\hat{\alpha}_1$ is

$$\text{MSE}\{\hat{\alpha}_1\} = \left(\tilde{\boldsymbol{a}}_1^H \tilde{\boldsymbol{a}}_1\right)^{-1} \tilde{\boldsymbol{a}}_1^H \left(\boldsymbol{\Xi}\boldsymbol{\beta}\boldsymbol{\beta}^H\boldsymbol{\Xi}^H + \boldsymbol{\Gamma}\right) \tilde{\boldsymbol{a}}_1 \left(\tilde{\boldsymbol{a}}_1^H \tilde{\boldsymbol{a}}_1\right)^{-1}. \tag{13}$$

Similar to LSE$(1,0,K)$, LSE$(1,0,1)$ is statistically inefficient for colored noise, but is asymptotically statistically efficient:

Corollary 1 *Under the same conditions as stated in Theorem 1, LSE*$(1,0,1)$ *is asymptotically statistically efficient, and the asymptotic MSE of* $\hat{\alpha}_1$ *is given by*

$$\lim_{N,\bar{N}\to\infty} N\bar{N}\textit{MSE}\{\hat{\alpha}_1\} = \lim_{N,\bar{N}\to\infty} N\bar{N}\textit{CRB}\{\alpha_1\} = \phi(f_1, \bar{f}_1). \tag{14}$$

Proof: Corollary 1 is easily proved by using (50) and the fact that $\lim_{N,\bar{N}\to\infty}(N\bar{N})^{-1/2}\tilde{\boldsymbol{a}}_1^H\boldsymbol{\Xi} = 0$ [cf. (48)]. □

Although the results in Theorem 1 and Corollary 1 indicate that LSE$(1,0,K)$ and LSE$(1,0,1)$ are asymptotically equivalent, they behave quite differently in finite samples. For most cases of interest, LSE$(1,0,K)$ is in general a better estimator than LSE$(1,0,1)$. Yet, the latter is usually preferred when we have a relatively large number of data samples, due to its computational simplicity.

3.2 Weighted-Least-Squares Amplitude Estimators

The LS estimators completely ignore the correlation of noise samples and, hence, are suboptimal in general. One way to take the noise correlation into consideration and, thus, to improve the estimation accuracy is to partition the observation matrix into overlapping submatrices. Doing so makes it possible to estimate the noise correlation which can be used to obtain a Markov-like weighted-least-squares (WLS) based estimator [13]. In the sequel, we discuss various ways to estimate the noise correlation, which result in several WLS amplitude estimators.

Let

$$\mathbf{X}_{l,\bar{l}} = \begin{bmatrix} x(l,\bar{l}) & \cdots & x(l,\bar{l}+\bar{M}-1) \\ x(l+1,\bar{l}) & \cdots & x(l+1,\bar{l}+\bar{M}-1) \\ \vdots & \vdots & \vdots \\ x(l+M-1,\bar{l}) & \cdots & x(l+M-1,\bar{l}+\bar{M}-1) \end{bmatrix}, \tag{15}$$
$$l = 0,\cdots,L-1;\ \bar{l} = 0,\cdots,\bar{L}-1,$$

where

$$L \triangleq N - M + 1, \quad \bar{L} \triangleq \bar{N} - \bar{M} + 1. \tag{16}$$

Let $\mathbf{V}_{l,\bar{l}} \in \mathbb{C}^{M\times\bar{M}}$ be similarly formed from $\{v(n,\bar{n})\}$. The choices of M and $\bar{M}$ are discussed in Section 4.2. Similar to (3), $\mathbf{X}_{l,\bar{l}}$ can be expressed as

$$\mathbf{X}_{l,\bar{l}} = \mathbf{A}_M \mathbf{\Lambda}_{l,\bar{l}} \bar{\mathbf{A}}_{\bar{M}}^T + \mathbf{V}_{l,\bar{l}}, \tag{17}$$

where

$$\mathbf{A}_M \stackrel{\triangle}{=} \left[\boldsymbol{a}_M(f_1) \quad \cdots \quad \boldsymbol{a}_M(f_K)\right], \quad \bar{\mathbf{A}}_{\bar{M}} \stackrel{\triangle}{=} \left[\bar{\boldsymbol{a}}_{\bar{M}}(\bar{f}_1) \quad \cdots \quad \bar{\boldsymbol{a}}_{\bar{M}}(\bar{f}_K)\right],$$
$$\boldsymbol{a}_M(f_k) \stackrel{\triangle}{=} \left[1 \quad \cdots \quad e^{j(M-1)2\pi f_k}\right]^T, \quad \bar{\boldsymbol{a}}_{\bar{M}}(\bar{f}_k) \stackrel{\triangle}{=} \left[1 \quad \cdots \quad e^{j(\bar{M}-1)2\pi \bar{f}_k}\right]^T,$$
$$\mathbf{\Lambda}_{l,\bar{l}} \stackrel{\triangle}{=} \mathbf{\Lambda}\,\mathrm{diag}\left\{e^{j2\pi(f_1 l+\bar{f}_1\bar{l})} \quad \cdots \quad e^{j2\pi(f_K l+\bar{f}_K\bar{l})}\right\} \stackrel{\triangle}{=} \mathbf{\Lambda}\mathbf{\Omega}_{l,\bar{l}}.$$

Let $\boldsymbol{x}_{l,\bar{l}} = \mathrm{vec}\{\mathbf{X}_{l,\bar{l}}\}$ and $\boldsymbol{v}_{l,\bar{l}} = \mathrm{vec}\{\mathbf{V}_{l,\bar{l}}\}$. Vectorizing both sides of (17) yields

$$\boldsymbol{x}_{l,\bar{l}} = (\bar{\mathbf{A}}_{\bar{M}} \otimes \mathbf{A}_M)\mathrm{vec}\{\mathbf{\Lambda}_{l,\bar{l}}\} + \boldsymbol{v}_{l,\bar{l}} \stackrel{\triangle}{=} \mathbf{\Psi}_{M\bar{M}}\mathbf{\Omega}_{l,\bar{l}}\boldsymbol{\alpha} + \boldsymbol{v}_{l,\bar{l}}. \tag{18}$$

where [cf. (6)]

$$\mathbf{\Psi}_{M\bar{M}} \stackrel{\triangle}{=} \left[\bar{\boldsymbol{a}}_{\bar{M}}(\bar{f}_1)\otimes\boldsymbol{a}_M(f_1) \quad \ldots \quad \bar{\boldsymbol{a}}_{\bar{M}}(\bar{f}_K)\otimes\boldsymbol{a}_M(f_K)\right].$$

3.2.1 WLSE$(L\bar{L}, 0, K)$

The WLS (Markov-like) estimate of $\boldsymbol{\alpha}$ in (18) is given by [13]

$$\hat{\boldsymbol{\alpha}} = \left(\sum_{l=0}^{L-1}\sum_{\bar{l}=0}^{\bar{L}-1} \mathbf{\Omega}_{l,\bar{l}}^H \mathbf{\Psi}_{M\bar{M}}^H \hat{\mathbf{Q}}^{-1} \mathbf{\Psi}_{M\bar{M}} \mathbf{\Omega}_{l,\bar{l}}\right)^{-1} \left(\sum_{l=0}^{L-1}\sum_{\bar{l}=0}^{\bar{L}-1} \mathbf{\Omega}_{l,\bar{l}}^H \mathbf{\Psi}_{M\bar{M}}^H \hat{\mathbf{Q}}^{-1} \boldsymbol{x}_{l,\bar{l}}\right), \tag{19}$$

where $\hat{\mathbf{Q}}$ is an estimate of $\mathbf{Q} \stackrel{\triangle}{=} E\{\boldsymbol{v}_{l,\bar{l}}\boldsymbol{v}_{l,\bar{l}}^H\}$.

An estimate of $\mathbf{Q}$ may be obtained as follows. Assume that the initial phases of the sinusoids are independently, identically and uniformly distributed over $[-\pi, \pi)$, and are independent of the noise in (18). Then we have

$$\mathbf{R} \stackrel{\triangle}{=} E\{\boldsymbol{x}_{l,\bar{l}}\boldsymbol{x}_{l,\bar{l}}^H\} = \mathbf{\Psi}_{M\bar{M}}\mathbf{P}\mathbf{\Psi}_{M\bar{M}}^H + \mathbf{Q}, \tag{20}$$

where $\mathbf{P} = \mathrm{diag}\left\{|\alpha_1^2| \quad \cdots \quad |\alpha_K^2|\right\}$. A straightforward estimate of $\mathbf{Q}$ is then obtained as

$$\hat{\mathbf{Q}} = \hat{\mathbf{R}} - \mathbf{\Psi}_{M\bar{M}}\hat{\mathbf{P}}\mathbf{\Psi}_{M\bar{M}}^H, \tag{21}$$

where $\hat{\mathbf{P}}$ is some initial estimate of $\mathbf{P}$, and $\hat{\mathbf{R}}$ is the sample covariance matrix of $\{\boldsymbol{x}_{l,\bar{l}}\}$, i.e.,

$$\hat{\mathbf{R}} \stackrel{\triangle}{=} \frac{1}{L\bar{L}}\sum_{l=0}^{L-1}\sum_{\bar{l}=0}^{\bar{L}-1} \boldsymbol{x}_{l,\bar{l}}\boldsymbol{x}_{l,\bar{l}}^H. \tag{22}$$

In order to eliminate the need for an initial estimate $\hat{\mathbf{P}}$, we observe that

$$\hat{\mathbf{R}}\hat{\mathbf{Q}}^{-1}\mathbf{\Psi}_{M\bar{M}} = \mathbf{\Psi}_{M\bar{M}}\left(\hat{\mathbf{P}}\mathbf{\Psi}_{M\bar{M}}^H\hat{\mathbf{Q}}^{-1}\mathbf{\Psi}_{M\bar{M}} + \mathbf{I}_K\right).$$

For sufficiently large $N, \bar{N}, M$, and $\bar{M}$, $\mathbf{\Psi}_{M\bar{M}}^H\hat{\mathbf{Q}}^{-1}\mathbf{\Psi}_{M\bar{M}}$ is approximately a diagonal matrix [cf. (52)] and, hence, so is $\mathbf{\Sigma} \stackrel{\triangle}{=} \hat{\mathbf{P}}\mathbf{\Psi}_{M\bar{M}}^H\hat{\mathbf{Q}}^{-1}\mathbf{\Psi}_{M\bar{M}} + \mathbf{I}_K$. Therefore,

$$\hat{\mathbf{Q}}^{-1}\mathbf{\Psi}_{M\bar{M}}\mathbf{\Omega}_{l,\bar{l}} = \hat{\mathbf{R}}^{-1}\mathbf{\Psi}_{M\bar{M}}\mathbf{\Sigma}\mathbf{\Omega}_{l,\bar{l}} \approx \hat{\mathbf{R}}^{-1}\mathbf{\Psi}_{M\bar{M}}\mathbf{\Omega}_{l,\bar{l}}\mathbf{\Sigma}. \tag{23}$$

It follows from (19) and (23) that

$$\hat{\boldsymbol{\alpha}} = \left(\sum_{l=0}^{L-1}\sum_{\bar{l}=0}^{\bar{L}-1}\mathbf{\Omega}_{l,\bar{l}}^H\mathbf{\Psi}_{M\bar{M}}^H\hat{\mathbf{R}}^{-1}\mathbf{\Psi}_{M\bar{M}}\mathbf{\Omega}_{l,\bar{l}}\right)^{-1}\left(\sum_{l=0}^{L-1}\sum_{\bar{l}=0}^{\bar{L}-1}\mathbf{\Omega}_{l,\bar{l}}^H\mathbf{\Psi}_{M\bar{M}}^H\hat{\mathbf{R}}^{-1}\boldsymbol{x}_{l,\bar{l}}\right). \tag{24}$$

The amplitude estimator (24) is an extension of the 2-D Capon spectral estimator (see, e.g., [4], [9]) to multiple sinusoids.

An alternative estimate of $\mathbf{Q}$, other than (21), can be obtained as described next. Let

$$\tilde{\boldsymbol{a}}_{M\bar{M}}(f_k, \bar{f}_k) \stackrel{\triangle}{=} \bar{\boldsymbol{a}}_{\bar{M}}(\bar{f}_k) \otimes \boldsymbol{a}_M(f_k).$$

Then (18) can be expressed as

$$\boldsymbol{x}_{l,\bar{l}} = \sum_{k=1}^{K}\alpha_k\tilde{\boldsymbol{a}}_{M\bar{M}}(f_k, \bar{f}_k)e^{j2\pi(f_k l + \bar{f}_k\bar{l})} + \boldsymbol{v}_{l,\bar{l}}, \quad l = 0, \cdots, L-1;\ \bar{l} = 0, \cdots, \bar{L}-1. \tag{25}$$

The unstructured LS estimate of $\alpha_k\tilde{\boldsymbol{a}}_{M\bar{M}}(f_k, \bar{f}_k)$ from (25) is given by

$$\widehat{\alpha_k\tilde{\boldsymbol{a}}_{M\bar{M}}(f_k, \bar{f}_k)} = \frac{1}{L\bar{L}}\sum_{l=0}^{L-1}\sum_{\bar{l}=0}^{\bar{L}-1}\boldsymbol{x}_{l,\bar{l}}e^{-j2\pi(f_k l + \bar{f}_k\bar{l})} \stackrel{\triangle}{=} \boldsymbol{\xi}_k, \tag{26}$$

which is the normalized 2-D DFT of the vector sequence $\{\boldsymbol{x}_{l,\bar{l}}\}$. Note that

$$\mathbf{\Psi}_{M\bar{M}}\mathbf{P}\mathbf{\Psi}_{M\bar{M}}^H = \sum_{k=1}^{K}\left[\alpha_k\tilde{\boldsymbol{a}}_{M\bar{M}}(f_k, \bar{f}_k)\right]\left[\alpha_k\tilde{\boldsymbol{a}}_{M\bar{M}}(f_k, \bar{f}_k)\right]^H. \tag{27}$$

It follows from (20) and (27) that a new estimate of $\mathbf{Q}$, which is different from (21), can be obtained as

$$\hat{\mathbf{Q}} = \hat{\mathbf{R}} - \sum_{k=1}^{K}\left[\alpha_k\widehat{\tilde{\boldsymbol{a}}_{M\bar{M}}(f_k, \bar{f}_k)}\right]\left[\alpha_k\widehat{\tilde{\boldsymbol{a}}_{M\bar{M}}(f_k, \bar{f}_k)}\right]^H = \hat{\mathbf{R}} - \sum_{k=1}^{K}\boldsymbol{\xi}_k\boldsymbol{\xi}_k^H. \tag{28}$$

The WLSE$(L\bar{L}, 0, K)$ that uses (19) with $\hat{\mathbf{Q}}$ given in (28) does not require any initial amplitude estimates. It is an extension of the 2-D APES algorithm [4], [9] to multiple sinusoids with known frequencies.

3.2.2 WLSE$(L\bar{L}, 0, 1)$

If we apply the WLS technique as in the previous section but restrict it to estimate one sinusoid at a time, then the WLSE$(L\bar{L}, 0, K)$ amplitude estimator reduces to WLSE$(L\bar{L}, 0, 1)$. In particular, the WLSE$(L\bar{L}, 0, 1)$ estimator that corresponds to using (19) with $\hat{\mathbf{Q}}$ given by (21) reduces to

$$\hat{\alpha}_k = \frac{\tilde{\boldsymbol{a}}^H_{M\bar{M}}(f_k, \bar{f}_k)\hat{\mathbf{R}}^{-1}\boldsymbol{\xi}_k}{\tilde{\boldsymbol{a}}^H_{M,\bar{M}}(f_k, \bar{f}_k)\hat{\mathbf{R}}^{-1}\tilde{\boldsymbol{a}}_{M,\bar{M}}(f_k, \bar{f}_k)}, \quad k = 1, \cdots, K. \tag{29}$$

On the other hand, the WLSE$(L\bar{L}, 0, 1)$ estimator that corresponds to using (19) with $\hat{\mathbf{Q}}$ as in (28) is given by

$$\hat{\alpha}_k = \frac{\tilde{\boldsymbol{a}}^H_{M\bar{M}}(f_k, \bar{f}_k)\big(\hat{\mathbf{R}} - \boldsymbol{\xi}_k\boldsymbol{\xi}_k^H\big)^{-1}\boldsymbol{\xi}_k}{\tilde{\boldsymbol{a}}^H_{M,\bar{M}}(f_k, \bar{f}_k)\big(\hat{\mathbf{R}} - \boldsymbol{\xi}_k\boldsymbol{\xi}_k^H\big)^{-1}\tilde{\boldsymbol{a}}_{M,\bar{M}}(f_k, \bar{f}_k)}, \quad k = 1, \cdots, K. \tag{30}$$

It should be stressed that, unlike (24), (29) is *exactly* equivalent to using (21) with (19). The amplitude estimators (29) and (30) have the same form as the 2-D Capon and APES spectral estimators [4], [9]. It was shown in [9] that both estimators are asymptotically efficient, but with quite different finite-sample properties. In particular, (29) is *biased downward*, whereas (30) is unbiased within a second-order approximation.

3.3 Matched-Filterbank Amplitude Estimator

The essence of the MAtched-FIlterbank (MAFI) approach to amplitude estimation is to design a bank of frequency-selective FIR (finite impulse response) filters whose center frequencies correspond to the 2-D frequencies of the sinusoids. The observed data is passed through the FIR filters, and the amplitude estimates are then obtained from the filtered and, hopefully, signal enhanced data. The role of the FIR filters is to suppress the interference and noise from adjacent frequencies, in an effort to maximize the SNR at the frequencies of interest.

Let $\mathbf{H}^H \in \mathbb{C}^{K\times M\bar{M}}$ be such that each row of $\mathbf{H}^H$ corresponds to an $M\bar{M}$-tap FIR filter. It was shown in [1] that choosing more than K FIR filters will not improve the SNR at the filterbank output, but choosing less than K FIR filters will certainly decrease the SNR. Therefore, we consider a bank of K FIR filters, each centered to the 2-D frequencies of one sinusoid of interest. Applying $\mathbf{H}^H$ to both sides of (18), we obtain the filterbank output as

$$\boldsymbol{y}_{l,\bar{l}} \stackrel{\triangle}{=} \mathbf{H}^H\boldsymbol{x}_{l,\bar{l}} = \mathbf{H}^H\boldsymbol{\Psi}_{M\bar{M}}\boldsymbol{\Omega}_{l,\bar{l}}\boldsymbol{\alpha} + \mathbf{H}^H\boldsymbol{v}_{l,\bar{l}}. \tag{31}$$

The FIR filters $\mathbf{H}^H$ are designed to maximize the SNR at the filterbank output:

$$\text{SNR} \stackrel{\triangle}{=} \text{tr}\left\{(\mathbf{H}^H\hat{\mathbf{Q}}\mathbf{H})^{-1}\mathbf{H}^H(\boldsymbol{\Psi}_{M\bar{M}}\hat{\mathbf{P}}\boldsymbol{\Psi}^H_{M\bar{M}})\mathbf{H}\right\},$$

that is,

$$\mathbf{H} = \arg \max_{\mathbf{H} \in \mathbb{C}^{M\bar{M} \times K}} \operatorname{tr} \left\{ (\mathbf{H}^H \hat{\mathbf{Q}} \mathbf{H})^{-1} \mathbf{H}^H (\mathbf{\Psi}_{M\bar{M}} \hat{\mathbf{P}} \mathbf{\Psi}_{M\bar{M}}^H) \mathbf{H} \right\}, \tag{32}$$

where $\operatorname{tr}\{\cdot\}$ denotes the trace of a matrix. The solution to (32) is *not unique* [1]. One that has a simple closed-form is given by [1]

$$\mathbf{H} = \hat{\mathbf{Q}}^{-1} \mathbf{\Psi}_{M\bar{M}} (\mathbf{\Psi}_{M\bar{M}}^H \hat{\mathbf{Q}}^{-1} \mathbf{\Psi}_{M\bar{M}})^{-1}. \tag{33}$$

It is easily verified that the above $\mathbf{H}$ satisfies

$$\mathbf{H}^H \mathbf{\Psi}_{M\bar{M}} = \mathbf{I}_K,$$

which implies that each FIR filter (row) in $\mathbf{H}^H$ passes the sinusoid of interest undistorted (with unit gain) and completely eliminates the other sinusoids. Therefore, (31) can be written as

$$\boldsymbol{y}_{l,\bar{l}} = \mathbf{\Omega}_{l,\bar{l}} \boldsymbol{\alpha} + \mathbf{H}^H \boldsymbol{v}_{l,\bar{l}}. \tag{34}$$

Observe that the covariance matrix of the noise term, $\mathbf{H}^H \boldsymbol{v}_{l,\bar{l}}$, in (34) is given by

$$\mathbf{H}^H \hat{\mathbf{Q}} \mathbf{H} = (\mathbf{\Psi}_{M\bar{M}}^H \hat{\mathbf{Q}}^{-1} \mathbf{\Psi}_{M\bar{M}})^{-1} \triangleq \mathbf{\Theta}.$$

Applying the WLS (Markov-like) technique to (34), we obtain the MAFI amplitude estimate

$$\begin{aligned}
\hat{\boldsymbol{\alpha}} =& \left(\sum_{l=0}^{L-1} \sum_{\bar{l}=0}^{\bar{L}-1} \mathbf{\Omega}_{l,\bar{l}}^H \mathbf{\Theta}^{-1} \mathbf{\Omega}_{l,\bar{l}} \right)^{-1} \left(\sum_{l=0}^{L-1} \sum_{\bar{l}=0}^{\bar{L}-1} \mathbf{\Omega}_{l,\bar{l}}^H \mathbf{\Theta}^{-1} \boldsymbol{y}_{l,\bar{l}} \right) \\
=& \left(\sum_{l=0}^{L-1} \sum_{\bar{l}=0}^{\bar{L}-1} \mathbf{\Omega}_{l,\bar{l}}^H \mathbf{\Psi}_{M\bar{M}}^H \hat{\mathbf{Q}}^{-1} \mathbf{\Psi}_{M\bar{M}} \mathbf{\Omega}_{l,\bar{l}} \right)^{-1} \\
& \times \left(\sum_{l=0}^{L-1} \sum_{\bar{l}=0}^{\bar{L}-1} \mathbf{\Omega}_{l,\bar{l}}^H \left[\mathbf{\Psi}_{M\bar{M}}^H \hat{\mathbf{Q}}^{-1} \mathbf{\Psi}_{M\bar{M}} \right] \left[\mathbf{\Psi}_{M\bar{M}}^H \hat{\mathbf{Q}}^{-1} \mathbf{\Psi}_{M\bar{M}} \right]^{-1} \mathbf{\Psi}_{M\bar{M}}^H \hat{\mathbf{Q}}^{-1} \boldsymbol{x}_{l,\bar{l}} \right) \\
=& \left(\sum_{l=0}^{L-1} \sum_{\bar{l}=0}^{\bar{L}-1} \mathbf{\Omega}_{l,\bar{l}}^H \mathbf{\Psi}_{M\bar{M}}^H \hat{\mathbf{Q}}^{-1} \mathbf{\Psi}_{M\bar{M}} \mathbf{\Omega}_{l,\bar{l}} \right)^{-1} \left(\sum_{l=0}^{L-1} \sum_{\bar{l}=0}^{\bar{L}-1} \mathbf{\Omega}_{l,\bar{l}}^H \mathbf{\Psi}_{M\bar{M}}^H \hat{\mathbf{Q}}^{-1} \boldsymbol{x}_{l,\bar{l}} \right),
\end{aligned} \tag{35}$$

where in the second equality, we used the fact that $\boldsymbol{y}_{l,\bar{l}} = \mathbf{H}^H \boldsymbol{x}_{l,\bar{l}}$ [cf. (31)] with $\mathbf{H}$ given by (33). Hence, we have MAFI$(L\bar{L}, K, K)$ = WLSE$(L\bar{L}, 0, K)$. Similar to the 1-D case in [1], the MAFI$(L\bar{L}, K, K)$ approach is more general since it includes WLSE$(L\bar{L}, 0, K)$ as a special case and other MAFI estimators which do not have a WLS interpretation exist. Next, we derive such a MAFI estimator.

Let $y_{l,\bar{l}}(k)$ and $\nu_{l,\bar{l}}(k)$ be the k-th element of $\boldsymbol{y}_{l,\bar{l}}$ and $\mathbf{H}^H \boldsymbol{v}_{l,\bar{l}}$, respectively. Then (34) can be written as

$$y_{l,\bar{l}}(k) = \alpha_k e^{j2\pi(f_k l + \bar{f}_k \bar{l})} + \nu_{l,\bar{l}}(k), \quad k = 1, \ldots, K. \tag{36}$$

Applying LS to the above equation gives the MAFI$(L\bar{L}, K, 1)$ estimator:

$$\hat{\alpha}_k = \frac{1}{L\bar{L}} \sum_{l=0}^{L-1} \sum_{\bar{l}=0}^{\bar{L}-1} y_{l,\bar{l}}(k) e^{-j2\pi(f_k l + \bar{f}_k \bar{l})}, \quad k = 1, \cdots, K. \tag{37}$$

Unlike the other one-at-a-time estimators, the above MAFI$(L\bar{L}, K, 1)$ estimator requires knowledge of the number and frequencies of the sinusoids, owing to the need to design the filterbank.

4 Numerical Examples

In this section, we evaluate the amplitude estimators discussed in the previous section. Since all estimators are asymptotically efficient and, thus, asymptotically equivalent, we only consider cases when N and $\bar{N}$ are relatively small. The following acronyms are used to distinguish the different estimators:

- LSE1: LSE$(1, 0, 1)$ using (11);
- LSEK: LSE$(1, 0, K)$ using (7);
- Capon1: WLSE$(L\bar{L}, 0, 1)$ using (29);
- APES1: WLSE$(L\bar{L}, 0, 1)$ using (30);
- CaponK: WLSE$(L\bar{L}, 0, K)$ using (24);
- APESK: WLSE$(L\bar{L}, 0, K)$ using (19) along with (28);
- MAFI1: MAFI$(L\bar{L}, K, 1)$ using (37) along with (28).

The data consists of $K = 3$ 2-D complex sinusoids contaminated by a zero-mean complex Gaussian noise. The frequencies of the sinusoids are $(0.45, 0.35)$, $(0.235, 0.135)$, and $(0.2, 0.1)$, respectively, whereby the second and third sinusoids are close to each other but the first one is away from the other two. The amplitudes of the three sinusoids are $\alpha_1 = e^{j\pi/4}$, $\alpha_2 = e^{j\pi/3}$, and $\alpha_3 = e^{j\pi/4}$, respectively. All performances are evaluated based on 200 independent Monte Carlo realizations.

4.1 Performance versus SNR

We begin by considering the case when the noise is colored. Specifically, the colored noise is generated by a 2-D autoregressive (AR) process:

$$v(n, \bar{n}) = 0.99 v(n-1, \bar{n}-1) + e(n, \bar{n}), \quad n = 0, \cdots, N-1; \ \bar{n} = 0, \cdots, \bar{N}-1, \tag{38}$$

where $e(n, \bar{n})$ is a complex-valued white Gaussian noise with zero-mean and variance σ^2. The power spectral density (PSD) of the data is shown in Figure 1 where $\sigma^2 = 0.01$. The SNR of the k-th sinusoid is defined as [2]

$$\text{SNR}_k = 10 \log_{10} \frac{|\alpha_k|^2}{\phi(f_k, \bar{f}_k)}.$$

In this example, we have 16×16 data samples, that is, $N = \bar{N} = 16$. For the WLS and MAFI estimators, we choose $M = \bar{M} = 4$.

Figure 2(a) shows the mean squared errors (MSE) of the seven estimators for α_1, along with the CRB given in (9), as the SNR changes. We see that APES1, APESK, and MAFI1 are close to the CRB, while both LS estimators are away from the CRB. CaponK is away from the CRB at high SNRs, which is due to a bias introduced in the approximation of (23) [1]. Capon1 is also biased for finite samples [9], which causes it to deviate from the CRB at high SNRs.

The performances of the above estimators are somewhat different when some sinusoids are close to the one of interest, as in the case shown in Figure 2(b), which gives the MSEs and CRBs corresponding to α_3. In particular, LSE1, Capon1, and APES1, which all estimate one sinusoid at a time, degrade considerably. Unlike LSE1 and Capon1, APES1 still appears to be consistent in that its MSE continuously decreases as the SNR increases. The performance of APESK becomes quite sensitive for certain SNRs, as also observed in the 1-D scenarios [1]. In the current case, MAFI1 appears to be the best estimator.

As shown in Section 3.1, when the noise is white and $N\bar{N} \geq K$, LSEK achieves the CRB and, thus, is optimal. It is interesting to know how the other suboptimal estimators perform in such a case. We consider an example similar to the previous one except that $v(n, \bar{n})$ is now a zero-mean complex white Gaussian noise. The results are shown in Figures 3(a) and 3(b). For α_1, all estimators perform quite well and are close to the CRB for most SNRs. When the SNR is high, however, the bias of LSE1, Capon1, and CaponK dominates their MSE, which causes their performance deviation away from the CRB. For α_3, all suboptimal estimators are observed to degrade to various extents. Yet, the difference between MAFI1 and LSEK is still relatively small.

Summary of 2-D Amplitude Estimators: APES1 is in general preferred when no sinusoids are closely spaced, or when the closely spaced sinusoids are of no interest to the study. In such a case, APES1 is preferred over APESK or MAFI1 since it is computationally simpler and, moreover, it does not require the knowledge of the sinusoidal frequencies. On the other hand, MAFI1 should be preferred when estimating closely spaced sinusoids. Capon1 is generally not recommended due to its bias. LSEK may be used when it is known *a priori* that the observation noise is white. Finally, LSE1 may be preferred when the number of data samples is large, in which case all estimators are similar to one another but LSE1 is computationally

most efficient.

4.2 Effect of M and $\bar{M}$ on the Performances

We now study the effect of the filter size parameters M and $\bar{M}$ on the WLS and MAFI estimators, whose performance depends on these parameters. The test data is generated as described in the previous section, with the observation noise being colored and $N = \bar{N} = 32$. Without loss of generality, we choose $M = \bar{M}$. The SNR is fixed at 20 dB as M and $\bar{M}$ vary from 2 to 16. It should be noted from (22) that we need $M\bar{M} \leq L\bar{L}$ in order to ensure that the sample covariance $\hat{\mathbf{R}}$ has full rank.

The performance of the estimators in terms of MSE and the corresponding CRB versus $M = \bar{M}$ for α_1 and α_3 is shown in Figures 4(a) and 4(b), respectively. When no sinusoids are close to the one of interest, as in the case shown in Figure 4(a) for α_1, only Capon1 and CaponK are rather sensitive to the choice of M and $\bar{M}$; the performance of the other WLS and MAFI estimators remains relatively unaffected for a wide range of M and $\bar{M}$. However, when some sinusoids are close to the one of interest, as in the case shown in Figure 4(b) for α_3, APESK and MAFI1 become more sensitive to the choice of M and $\bar{M}$, whereas APES1 still keeps its insensitivity to the different choices of M and $\bar{M}$. Hence, choosing the right filter size for APES1 is easier than for the other WLS and MAFI estimators.

5 Application to System Identification

In this section, we use the 2-D amplitude estimation techniques to solve a problem of 2-D system identification. Consider the following 2-D linear discrete-time system

$$x(n,\bar{n}) = H(z^{-1},\bar{z}^{-1})u(n,\bar{n}) + v(n,\bar{n}), \quad n = 0,\cdots,N-1;\ \bar{n} = 0,\cdots,\bar{N}-1, \tag{39}$$

where the input, also called the "probing signal", is a summation of K 2-D complex-valued sinusoids:

$$u(n,\bar{n}) = \sum_{k=1}^{K} \gamma_k e^{j2\pi(f_k n + \bar{f}_k \bar{n})}, \tag{40}$$

$v(n,\bar{n})$ is the measurement noise, and the system transfer function $H(z^{-1},\bar{z}^{-1})$ is given by

$$H(z^{-1},\bar{z}^{-1}) = \frac{B(z^{-1},\bar{z}^{-1})}{A(z^{-1},\bar{z}^{-1})} = \frac{\sum_{i=0}^{r-1}\sum_{j=0}^{s-1} b_{i,j} z^{-i}\bar{z}^{-j}}{\sum_{i=0}^{p-1}\sum_{j=0}^{q-1} a_{i,j} z^{-i}\bar{z}^{-j}}, \tag{41}$$

where, without loss of generality, $a_{0,0} = 1$, $b_{0,0} = 0$, and $(z^{-1},\bar{z}^{-1})$ are the unit-delay operators: $z^{-1}\bar{z}^{-1}u(n,\bar{n}) = u(n-1,\bar{n}-1)$ [13]. We assume that the system orders $r,\ s,\ p$, and q are known, and that the number of

sinusoids, K, is no less than the number of unknown parameters, i.e., $K \geq (pq-1)+(rs-1)$. Let

$$\boldsymbol{a} \stackrel{\triangle}{=} \begin{bmatrix} a_{0,1} & \cdots & a_{0,q-1} & \cdots & a_{p-1,0} & \cdots & a_{p-1,q-1} \end{bmatrix}^T,$$
$$\boldsymbol{b} \stackrel{\triangle}{=} \begin{bmatrix} b_{0,1} & \cdots & b_{0,s-1} & \cdots & b_{r-1,0} & \cdots & b_{r-1,s-1} \end{bmatrix}^T.$$

The 2-D system identification problem of interest is to estimate the system parameters $\boldsymbol{a}$ and $\boldsymbol{b}$ from the system outputs $\{x(n,\bar{n})\}$.

5.1 System Identification Using Amplitude Estimation

A standard approach to solving the above 2-D system identification problem is the output error method (OEM), which minimizes the following cost function [13]:

$$C_{\text{OEM}}(\boldsymbol{a},\boldsymbol{b}) = \sum_{n=0}^{N-1}\sum_{\bar{n}=0}^{\bar{N}-1} \left| x(n,\bar{n}) - H(z^{-1},\bar{z}^{-1})u(n,\bar{n}) \right|^2. \tag{42}$$

The OEM cost function can be minimized by using standard iterative nonlinear search schemes, which are usually computationally intensive and can at best converge to a local minima of the OEM cost function.

In order to avoid the aforementioned drawback of OEM, we introduce a new 2-D system identification algorithm by making use of amplitude estimation techniques. This algorithm has a 1-D counterpart, originally proposed in [1] for 1-D system identification. Following the same argument as in the 1-D case, the new 2-D system identification algorithm presented in the sequel is asymptotically (for large sample) statistically efficient [1]. We sketch the derivation of this algorithm next. For more motivation and discussions of this technique, we refer the reader to [1].

Let

$$\alpha_k(\boldsymbol{a},\boldsymbol{b}) \stackrel{\triangle}{=} \gamma_k H(e^{-j2\pi f_k}, e^{-j2\pi \bar{f}_k}). \tag{43}$$

For sufficiently large N and $\bar{N}$ such that the transient response in the output can be ignored, (39) and (40) can be approximately expressed as

$$x(n,\bar{n}) = \sum_{k=1}^{K} \alpha_k(\boldsymbol{a},\boldsymbol{b}) e^{j2\pi(f_k n + \bar{f}_k \bar{n})} + v(n,\bar{n}), \quad n=0,\cdots,N-1;\ \bar{n}=0,\cdots,\bar{N}-1. \tag{44}$$

From the above equation, we can estimate $\{\alpha_k(\boldsymbol{a},\boldsymbol{b})\}_{k=1}^K$, in an *unstructured* manner, by using any amplitude estimator discussed in Section 3. Once the unstructured estimates of $\{\alpha_k(\boldsymbol{a},\boldsymbol{b})\}_{k=1}^K$ are obtained, we select the $(pq-1)+(rs-1)$ largest ones (in magnitude) out of the K amplitude estimates, and denote them by $\hat{\alpha}_k,\ k=1,\ldots,(pq-1)+(rs-1)$. Given these amplitude estimates, it is possible to choose $\boldsymbol{a}$ and $\boldsymbol{b}$ such that

$$\hat{\alpha}_k = \alpha_k(\boldsymbol{a},\boldsymbol{b}),\ k=1,\cdots,(pq-1)+(rs-1).$$

Equivalently, the above equation can be rewritten as

$$\hat{\alpha}_k A(e^{-j2\pi f_k}, e^{-j2\pi \bar{f}_k}) = \gamma_k B(e^{-j2\pi f_k}, e^{-j2\pi \bar{f}_k}), \quad k = 1, \cdots, (pq-1)+(rs-1), \tag{45}$$

which is a set of $(pq-1)+(rs-1)$ linear equations with $(pq-1)+(rs-1)$ unknowns. Solving these equations gives an estimate of the system parameters $\boldsymbol{a}$ and $\boldsymbol{b}$.

5.2 Numerical Results

We compare the performances of the OEM and the proposed 2-D system identification scheme. The OEM can be implemented by using any standard iterative nonlinear search algorithm. Here, we use `lsqnonlin` provided in MATLAB. The initial values of $\boldsymbol{a}$ and $\boldsymbol{b}$, which are required by `lsqnonlin`, are obtained by the LS approach [13, pp. 60–67]. In our simulations, we noticed that better performance can be achieved if $\boldsymbol{a}$ and $\boldsymbol{b}$ are estimated separately by using `lsqnonlin`. That is, we search for the minimizing $\boldsymbol{a}$ in the first step while fixing the value for $\boldsymbol{b}$; next, we minimize the OEM cost function with respect to $\boldsymbol{b}$ while fixing $\boldsymbol{a}$ to its most recently updated value. We iterate the two steps until practical convergence is attained.

Our new proposed scheme is non-iterative and in fact it has a closed-form. It can be implemented in various ways, depending on which amplitude estimator is used. In the following, we consider the LSEK, APES1, and MAFI1 amplitude estimators. Abusing the notations a little bit, we refer to the resulting system identification algorithms as LSEK, APES1, and MAFI1, respectively.

We first consider an example when the measurement noise, $v(n, \bar{n})$, is white. The system transfer function is described by

$$\begin{aligned} A(z^{-1}, \bar{z}^{-1}) &= 1 - j0.7\bar{z}^{-1} + (-0.529 - j0.7281)z^{-1} + (-0.5097 + j0.3703)z^{-1}\bar{z}^{-1}, \\ B(z^{-1}, \bar{z}^{-1}) &= (0.2163 - j0.6657)\bar{z}^{-1} + (-0.2472 - j0.7608)z^{-1} + (-0.56)z^{-1}\bar{z}^{-1}. \end{aligned}$$

The probing signal consists of $K = 8$ 2-D sinusoids at frequencies: $(-0.45,\ 0.48)$, $(-0.3167,\ 0.3467)$, $(-0.1833,\ 0.2133)$, $(-0.05,\ 0.08)$, $(0.05,\ -0.08)$, $(0.1833,\ -0.2133)$, $(0.3167,\ -0.3467)$, and $(0.45,\ -0.48)$. The noise $v(n, \bar{n})$ is complex Gaussian with zero-mean and variance $\sigma^2 = 0.01$. The performance criteria considered here are the averaged root mean squared errors (ARMSE) of the parameter estimates and the number of MATLAB flops associated with each method. The ARMSE for the $\boldsymbol{a}$ parameters is defined as

$$\text{ARMSE}\{\hat{\boldsymbol{a}}\} = \frac{1}{pq-1} \sum_{i,j,(i,j)\neq(0,0)} \text{RMSE}\{\hat{a}_{i,j}\}.$$

The ARMSE for the $\boldsymbol{b}$ parameters is similarly defined.

Figure 5 illustrates the performance of LSEK, APES1, MAFI1, and OEM, when $N = \bar{N}$ varies from 15 to 40, and $M = \bar{M} = 5$ for APES1 and MAFI1. In particular, Figures 5(a) and 5(b) show the ARMSE of the $\boldsymbol{a}$ and, respectively, $\boldsymbol{b}$ parameters, as well as the CRB that is derived in Appendix B. Figure 5(c) shows the number of flops required by each method. When the noise is white Gaussian, OEM is asymptotically optimal [13] and, therefore, it should outperform the proposed methods, as seen in Figures 5(a) and 5(b). However, the performance difference between OEM and the proposed methods, especially APES1 and MAFI1, is small. Figure 5(c) indicates that APES1 is slightly less involved than OEM. However, the differences in computational complexity are minor in this case. They will be more significant when the system becomes more complex or when the noise is colored, as will be seen in the next example.

Although LSEK is computationally the simplest, Figures 5(a) and 5(b) show that even with white noise, LSEK for system identification is considerably worse than APES1 and MAFI1. This may seem counter-intuitive since LSEK should obtain the best amplitude estimates when the observation noise is white (cf. Section 3.1.1). The poor performance of LSEK is due to the transient response of the 2-D system. In particular, (44) is no longer a good approximation of the observations for relatively small N and $\bar{N}$. Figures 6(a) and 6(b) show the power spectral density (PSD) estimate and, respectively, the contour plot of the PSD estimate of $x(n, \bar{n})$. The PSD estimate is obtained by the 2-D Capon PSD estimator [12], with $N = \bar{N} = 25$ and $M = \bar{M} = 8$. The system poles at $z = 0.9e^{j2\pi(0.15)}$ and $\bar{z} = 0.7e^{j2\pi(0.25)}$ generate substantial transient response in the system output, which manifests itself as a "ridge" along $f = 0.15$ and $\bar{f} = 0.25$, respectively. These ridges interfere with the 2-D sinusoids and lead to the degradation of LSEK.

Next, we consider an example that involves colored measurement noise. The system is given by

$$A(z^{-1}, \bar{z}^{-1}) = 1 - j0.8\bar{z}^{-1} - (0.529 + j0.7281)z^{-1} + (-0.5825 + j0.4232)z^{-1}\bar{z}^{-1},$$
$$B(z^{-1}, \bar{z}^{-1}) = (0.2014 - j0.7846)\bar{z}^{-1} - (0.2194 + j0.6753)z^{-1} + (-0.574 + j0.0361)z^{-1}\bar{z}^{-1}.$$

The probing signal is the same as in the previous example. The noise is generated by an AR process as in (38), with $e(n, \bar{n})$ being complex white Gaussian with zero-mean and variance $\sigma^2 = 0.01$. The results are shown in Figures 7(a) to 7(c). We see that in the current case, APES1 and MAFI1 not only obtain statistically more accurate parameter estimates than OEM does, but they are also computationally simpler than the latter.

6 Conclusions

In this paper, we have presented a number of amplitude estimators for 2-D sinusoidal signals in the presence of colored noise. Although all these amplitude estimators are asymptotically statistically efficient, they

have rather different finite-sample properties. The performances of these estimators in finite samples have been compared with one another and with the CRB. General guidelines for the selection and use of these estimators have been provided. As an application example, we have presented a new approach to 2-D system identification by making use of amplitude estimation techniques. It has been shown that when the measurement noise is colored, the new approach in general obtains better parameter estimates and is computationally simpler than the widely-used OEM algorithm.

Appendix A: Proof of Theorem 1

The MSE of $\hat{\boldsymbol{\alpha}}$ is given by (8), which is repeated below for easy reference:

$$\text{MSE}\{\hat{\boldsymbol{\alpha}}\} = (\boldsymbol{\Psi}^H\boldsymbol{\Psi})^{-1}\boldsymbol{\Psi}^H\boldsymbol{\Gamma}\boldsymbol{\Psi}(\boldsymbol{\Psi}^H\boldsymbol{\Psi})^{-1}. \tag{46}$$

In view of (6), we have

$$\boldsymbol{\Psi}^H\boldsymbol{\Psi} = \begin{bmatrix} (\bar{\boldsymbol{a}}_1^H\bar{\boldsymbol{a}}_1)(\boldsymbol{a}_1^H\boldsymbol{a}_1) & \dots & (\bar{\boldsymbol{a}}_1^H\bar{\boldsymbol{a}}_K)(\boldsymbol{a}_1^H\boldsymbol{a}_K) \\ \vdots & \vdots & \vdots \\ (\bar{\boldsymbol{a}}_K^H\bar{\boldsymbol{a}}_1)(\boldsymbol{a}_K^H\boldsymbol{a}_1) & \dots & (\bar{\boldsymbol{a}}_K^H\bar{\boldsymbol{a}}_K)(\boldsymbol{a}_K^H\boldsymbol{a}_K) \end{bmatrix} = (\bar{\mathbf{A}}^H\bar{\mathbf{A}}) \odot (\mathbf{A}^H\mathbf{A}), \tag{47}$$

where $\odot$ denotes the elementwise Hadamard product [14] and, for notational simplicity, $\boldsymbol{a}_k$ and $\bar{\boldsymbol{a}}_k$ stand for $\boldsymbol{a}(f_k)$ and $\bar{\boldsymbol{a}}(\bar{f}_k)$ [defined in (4)], respectively. Since $\lim_{N\to\infty}\frac{1}{N}(\mathbf{A}^H\mathbf{A}) = \mathbf{I}_K$ and $\lim_{\bar{N}\to\infty}\frac{1}{\bar{N}}(\bar{\mathbf{A}}^H\bar{\mathbf{A}}) = \mathbf{I}_K$ [15], it follows that

$$\lim_{N,\bar{N}\to\infty}\frac{1}{N\bar{N}}(\boldsymbol{\Psi}^H\boldsymbol{\Psi}) = \mathbf{I}_K. \tag{48}$$

In order to determine the asymptotic form of $\boldsymbol{\Psi}^H\boldsymbol{\Gamma}\boldsymbol{\Psi}$, we write it as

$$\boldsymbol{\Psi}^H\boldsymbol{\Gamma}\boldsymbol{\Psi} = \begin{bmatrix} (\bar{\boldsymbol{a}}_1 \otimes \boldsymbol{a}_1)^H\boldsymbol{\Gamma}(\bar{\boldsymbol{a}}_1 \otimes \boldsymbol{a}_1) & \dots & (\bar{\boldsymbol{a}}_1 \otimes \boldsymbol{a}_1)^H\boldsymbol{\Gamma}(\bar{\boldsymbol{a}}_K \otimes \boldsymbol{a}_K) \\ \vdots & \vdots & \vdots \\ (\bar{\boldsymbol{a}}_K \otimes \boldsymbol{a}_K)^H\boldsymbol{\Gamma}(\bar{\boldsymbol{a}}_1 \otimes \boldsymbol{a}_1) & \dots & (\bar{\boldsymbol{a}}_K \otimes \boldsymbol{a}_K)^H\boldsymbol{\Gamma}(\bar{\boldsymbol{a}}_K \otimes \boldsymbol{a}_K) \end{bmatrix}. \tag{49}$$

Since $\boldsymbol{\Gamma} = E\{\boldsymbol{v}\boldsymbol{v}^H\}$, the kl-th element of $\boldsymbol{\Psi}^H\boldsymbol{\Gamma}\boldsymbol{\Psi}$ can be expressed as

$$\begin{aligned} &(\bar{\boldsymbol{a}}_k \otimes \boldsymbol{a}_k)^H\boldsymbol{\Gamma}(\bar{\boldsymbol{a}}_l \otimes \boldsymbol{a}_l) \\ =&(\bar{\boldsymbol{a}}_k \otimes \boldsymbol{a}_k)^H E\{\boldsymbol{v}\boldsymbol{v}^H\}(\bar{\boldsymbol{a}}_l \otimes \boldsymbol{a}_l) \\ =&E\left\{\left[(\bar{\boldsymbol{a}}_k \otimes \boldsymbol{a}_k)^H\boldsymbol{v}\right]\left[\boldsymbol{v}^H(\bar{\boldsymbol{a}}_l \otimes \boldsymbol{a}_l)\right]\right\} \\ =&E\left\{V(f_k,\bar{f}_k)V^*(f_l,\bar{f}_l)\right\}, \end{aligned}$$

where $V(f_k,\bar{f}_k)$ is the 2-D discrete Fourier transform (DFT) of the noise $v(n,\bar{n})$ at the frequency pair $(f_k,\bar{f}_k)$, i.e.,

$$V(f_k,\bar{f}_k) = \sum_{n=0}^{N-1}\sum_{\bar{n}=0}^{\bar{N}-1} v(n,\bar{n})e^{-j2\pi(f_k n+\bar{f}_k\bar{n})}.$$

Hence, we have

$$\begin{aligned}
&\lim_{N,\bar{N}\to\infty} \frac{1}{N\bar{N}} \boldsymbol{\Psi}^H \boldsymbol{\Gamma} \boldsymbol{\Psi} \\
&= \lim_{N,\bar{N}\to\infty} \frac{1}{N\bar{N}} E \left\{ \begin{bmatrix} |V(f_1,\bar{f}_1)|^2 & \dots & V(f_1,\bar{f}_1)V^*(f_K,\bar{f}_K) \\ \vdots & \vdots & \vdots \\ V(f_K,\bar{f}_K)V^*(f_1,\bar{f}_1) & \dots & |V(f_K,\bar{f}_K)|^2 \end{bmatrix} \right\} \\
&= \operatorname{diag}\left\{ \phi(f_1,\bar{f}_1) \quad \dots \quad \phi(f_K,\bar{f}_K) \right\}.
\end{aligned} \tag{50}$$

where the last equality follows from the general definition of power spectral density (PSD) (see, e.g., [12]) and the fact that $\lim_{N,\bar{N}\to\infty} \frac{1}{N\bar{N}} E\{V(f_k,\bar{f}_k)V^*(f_l,\bar{f}_l)\} = 0$, for $k \neq l$ [15]. Therefore, the asymptotic MSE of $\hat{\boldsymbol{\alpha}}$ is [see (46), (48), and (50)]

$$\lim_{N,\bar{N}\to\infty} N\bar{N}\text{MSE}\{\hat{\boldsymbol{\alpha}}\} = \operatorname{diag}\left\{ \phi(f_1,\bar{f}_1) \quad \dots \quad \phi(f_K,\bar{f}_K) \right\}. \tag{51}$$

The Cramér-Rao bound (CRB) is given by (9). Similar to (50), it can be shown that [16]

$$\lim_{N,\bar{N}\to\infty} \frac{1}{N\bar{N}} \boldsymbol{\Psi}^H \boldsymbol{\Gamma}^{-1} \boldsymbol{\Psi} = \operatorname{diag}\left\{ \phi^{-1}(f_1,\bar{f}_1) \quad \dots \quad \phi^{-1}(f_K,\bar{f}_K) \right\}. \tag{52}$$

It follows from (9) and (52) that

$$\lim_{N,\bar{N}\to\infty} N\bar{N}\text{CRB}\{\boldsymbol{\alpha}\} = \operatorname{diag}\left\{ \phi(f_1,\bar{f}_1) \quad \dots \quad \phi(f_K,\bar{f}_K) \right\},$$

which coincides with (51) and concludes the proof.

Appendix B: Cramér-Rao Bound For The System Identification Problem

For the linear discrete-time system described by (39) and (41), we define $\boldsymbol{\theta}$ from the real and imaginary parts of the system coefficients:

$$\boldsymbol{\theta} \triangleq \begin{bmatrix} \boldsymbol{a}_R^T & \boldsymbol{b}_R^T & \boldsymbol{a}_I^T & \boldsymbol{b}_I^T \end{bmatrix}^T \in \mathbb{R}^{2[(pq-1)+(rs-1)]\times 1},$$

where we have

$$\begin{aligned}
\boldsymbol{a}_R &\triangleq \begin{bmatrix} \Re\{a_{0,1}\} & \cdots & \Re\{a_{0,q-1}\} & \cdots & \Re\{a_{p-1,0}\} & \cdots & \Re\{a_{p-1,q-1}\} \end{bmatrix}^T \in \mathbb{R}^{(pq-1)\times 1}, \\
\boldsymbol{b}_R &\triangleq \begin{bmatrix} \Re\{b_{0,1}\} & \cdots & \Re\{b_{0,s-1}\} & \cdots & \Re\{b_{r-1,0}\} & \cdots & \Re\{b_{r-1,s-1}\} \end{bmatrix}^T \in \mathbb{R}^{(rs-1)\times 1}, \\
\boldsymbol{a}_I &\triangleq \begin{bmatrix} \Im\{a_{0,1}\} & \cdots & \Im\{a_{0,q-1}\} & \cdots & \Im\{a_{p-1,0}\} & \cdots & \Im\{a_{p-1,q-1}\} \end{bmatrix}^T \in \mathbb{R}^{(pq-1)\times 1}, \\
\boldsymbol{b}_I &\triangleq \begin{bmatrix} \Im\{b_{0,1}\} & \cdots & \Im\{b_{0,s-1}\} & \cdots & \Im\{b_{r-1,0}\} & \cdots & \Im\{b_{r-1,s-1}\} \end{bmatrix}^T \in \mathbb{R}^{(rs-1)\times 1}.
\end{aligned} \tag{53}$$

Let $w(n,\bar{n}) \stackrel{\triangle}{=} H(z^{-1},\bar{z}^{-1})u(n,\bar{n})$. It is straightforward to verify that

$$w(n,\bar{n}) = \boldsymbol{\phi}_{n,\bar{n}}^T \boldsymbol{\theta}, \quad n = 0,\ldots,N-1;\ \bar{n} = 1,\ldots,\bar{N}-1,$$

where $\boldsymbol{\phi}_{n,\bar{n}} \in \mathbb{C}^{2[(pq-1)+(rs-1)]\times 1}$ is given by

$$\begin{aligned}\boldsymbol{\phi}_{n,\bar{n}} = \Big[& -w(n,\bar{n}-1) \quad \cdots \quad -w(n,\bar{n}-q+1) \quad \cdots \quad -w(n-p+1,\bar{n}) \quad \cdots \quad -w(n-p+1,\bar{n}-q+1) \\ & u(n,\bar{n}-1) \quad \cdots \quad u(n,\bar{n}-s+1) \quad \cdots \quad u(n-r+1,\bar{n}) \quad \cdots \quad u(n-r+1,\bar{n}-s+1) \\ & -jw(n,\bar{n}-1) \quad \cdots \quad -jw(n,\bar{n}-q+1) \quad \cdots \quad -jw(n-p+1,\bar{n}) \quad \cdots \quad -jw(n-p+1,\bar{n}-q+1) \\ & ju(n,\bar{n}-1) \quad \cdots \quad ju(n,\bar{n}-s+1) \quad \cdots \quad ju(n-r+1,\bar{n}) \quad \cdots \quad ju(n-r+1,\bar{n}-s+1) \Big]^T. \end{aligned} \tag{54}$$

Let $\mathbf{X} \in \mathbb{C}^{N\times\bar{N}}$, $\mathbf{W} \in \mathbb{C}^{N\times\bar{N}}$, and $\mathbf{V} \in \mathbb{C}^{N\times\bar{N}}$ be matrices formed from $\{x(n,\bar{n})\}$, $\{w(n,\bar{n})\}$, and $\{v(n,\bar{n})\}$, respectively, in the same manner as in (2). Let $\boldsymbol{x} \stackrel{\triangle}{=} \text{vec}\{\mathbf{X}\}$, $\boldsymbol{w} \stackrel{\triangle}{=} \text{vec}\{\mathbf{W}\}$, and $\boldsymbol{v} \stackrel{\triangle}{=} \text{vec}\{\mathbf{V}\}$. Then

$$\boldsymbol{x} = \boldsymbol{w} + \boldsymbol{v} = \boldsymbol{\Phi}\boldsymbol{\theta} + \boldsymbol{v},$$

where $\boldsymbol{\Phi} \in \mathbb{C}^{N\bar{N}\times 2[(pq-1)+(rs-1)]}$ is given by

$$\boldsymbol{\Phi} = \begin{bmatrix} \boldsymbol{\phi}_{0,0} & \boldsymbol{\phi}_{1,0} & \cdots & \boldsymbol{\phi}_{N-1,\bar{N}-1} \end{bmatrix}^T.$$

Under the assumption that the measurement noise $\boldsymbol{v}$ is circularly Gaussian with covariance matrix $\boldsymbol{\Gamma}$, the Cramér-Rao bound (CRB) for the system identification problem can be obtained by using the Slepian-Bangs formula [12]:

$$\text{CRB}^{-1}(\boldsymbol{\theta}) = 2\Re\left(\frac{\partial^H \boldsymbol{w}}{\partial \boldsymbol{\theta}} \boldsymbol{\Gamma}^{-1} \frac{\partial \boldsymbol{w}}{\partial \boldsymbol{\theta}^T} \right).$$

Next, we show how to obtain $\frac{\partial \boldsymbol{w}}{\partial \boldsymbol{\theta}^T}$. Let $a_{R_{i,j}} \stackrel{\triangle}{=} \text{Re}\{a_{i,j}\}$, $a_{I_{i,j}} \stackrel{\triangle}{=} \text{Im}\{a_{i,j}\}$, and let $b_{R_{i,j}}$ and $b_{I_{i,j}}$ be similarly defined from $b_{i,j}$. It is straightforward to verify that

$$\begin{aligned} \frac{\partial w(n,\bar{n})}{\partial a_{R_{i,j}}} &= -\frac{B(z^{-1},\bar{z}^{-1})}{A^2(z^{-1},\bar{z}^{-1})} u(n-i,\bar{n}-j) = -\frac{1}{A(z^{-1},\bar{z}^{-1})} w(n-i,\bar{n}-j), \\ \frac{\partial w(n,\bar{n})}{\partial b_{R_{i,j}}} &= \frac{1}{A(z^{-1},\bar{z}^{-1})} u(n-i,\bar{n}-j). \end{aligned}$$

Similarly, we have

$$\begin{aligned} \frac{\partial w(n,\bar{n})}{\partial a_{I_{i,j}}} &= -j\frac{1}{A(z^{-1},\bar{z}^{-1})} w(n-i,\bar{n}-j), \\ \frac{\partial w(n,\bar{n})}{\partial b_{I_{i,j}}} &= j\frac{1}{A(z^{-1},\bar{z}^{-1})} u(n-i,\bar{n}-j). \end{aligned}$$

It then follows from the definition of $\boldsymbol{\phi}_{n,\bar{n}}$ in (54) that

$$\frac{\partial w(n,\bar{n})}{\partial \boldsymbol{\theta}^T} = \frac{1}{A(z^{-1},\bar{z}^{-1})}\boldsymbol{\phi}_{n,\bar{n}}^T \stackrel{\triangle}{=} \boldsymbol{\zeta}_{n,\bar{n}}^T. \tag{55}$$

Let $\zeta_{n,\bar{n}}(i)$ and $\phi_{n,\bar{n}}(i)$ denote the i-th element of $\boldsymbol{\zeta}_{n,\bar{n}}$ and $\boldsymbol{\phi}_{n,\bar{n}}$, respectively. Equation (55) implies that $\zeta_{n,\bar{n}}(i)$ is obtained from $\phi_{n,\bar{n}}(i)$ through the linear regression:

$$A(z^{-1},\bar{z}^{-1})\zeta_{n,\bar{n}}(i) = \phi_{n,\bar{n}}(i), \quad i = 1,\cdots,2\left[(pq-1)+(rs-1)\right]. \tag{56}$$

Consequently, we have

$$\frac{\partial \boldsymbol{w}}{\partial \boldsymbol{\theta}^T} = \frac{1}{A(z^{-1},\bar{z}^{-1})}\boldsymbol{\Phi} \stackrel{\triangle}{=} \boldsymbol{\Delta},$$

where $\boldsymbol{\Delta} \in \mathbb{C}^{N\bar{N}\times 2[(pq-1)+(rs-1)]}$ with each element obtained in the same manner as in (56). It follows that the CRB can be written as

$$\text{CRB}(\boldsymbol{\theta}) = \frac{1}{2}\left[\Re\left(\boldsymbol{\Delta}^H\boldsymbol{\Gamma}^{-1}\boldsymbol{\Delta}\right)\right]^{-1}. \tag{57}$$

Note that $\boldsymbol{\Delta}$ has the form

$$\boldsymbol{\Delta} = \begin{bmatrix}\tilde{\boldsymbol{\Delta}} & j\tilde{\boldsymbol{\Delta}}\end{bmatrix},$$

where $\tilde{\boldsymbol{\Delta}}$ consists of the first $(pq-1)+(rs-1)$ columns of $\boldsymbol{\Delta}$. Then, it is readily verified that the CRB corresponding to the complex vector,

$$\tilde{\boldsymbol{\theta}} \stackrel{\triangle}{=} \begin{bmatrix}\boldsymbol{a}^T & \boldsymbol{b}^T\end{bmatrix}^T \in \mathbb{C}^{[(pq-1)+(rs-1)]\times 1},$$

is given by

$$\text{CRB}(\tilde{\boldsymbol{\theta}}) = \left(\tilde{\boldsymbol{\Delta}}^H\boldsymbol{\Gamma}^{-1}\tilde{\boldsymbol{\Delta}}\right)^{-1}. \tag{58}$$

Observe that the evaluation of the CRB (57) and (58) requires initial values of $w(n,\bar{n})$ for $1-p \le n \le -1$, $1-q \le \bar{n} \le -1$, and $u(n,\bar{n})$ for $1-r \le n \le -1$, $1-s \le \bar{n} \le -1$, which are set to zero in our simulations in Section 5.2.

References

[1] P. Stoica, H. Li, and J. Li, "Amplitude estimation of sinusoidal signals: Survey, new results, and an application," *IEEE Transactions on Signal Processing*, vol. 48, pp. 338–352, February 2000.

[2] S. M. Kay, *Modern Spectral Estimation: Theory and Application*, Prentice Hall, Englewood Cliffs, NJ, 1988.

[3] S. L. Marple, Jr., *Digital Spectral Analysis with Applications*, Prentice Hall, Englewood Cliffs, NJ, 1987.

[4] J. Li and P. Stoica, "An adaptive filtering approach to spectral estimation and SAR imaging," *IEEE Transactions on Signal Processing*, vol. 44, no. 6, pp. 1469–1484, June 1996.

[5] R. Agarwal, E. I. Plotkin, and M. N. S. Swamy, "Statistically optimal null filter based on instantaneous matched processing," *Circuits, Systems, and Signal Processing*, vol. 20, no. 1, pp. 38–61, January 2001.

[6] J. Capon, "High resolution frequency-wavenumber spectrum analysis," *Proceedings of the IEEE*, vol. 57, no. 8, pp. 1408–1418, August 1969.

[7] P. Stoica, H. Li, and J. Li, "A new derivation of the APES filter," *IEEE Signal Processing Letters*, vol. 6, no. 8, pp. 205–206, August 1999.

[8] P. Stoica, A. Jakobsson, and J. Li, "Matched-filter bank interpretation of some spectral estimators," *Signal Processing*, vol. 66, no. 1, pp. 45–59, April 1998.

[9] H. Li, J. Li, and P. Stoica, "Performance analysis of forward-backward matched-filterbank spectral estimators," *IEEE Transactions on Signal Processing*, vol. 46, no. 7, pp. 1954–1966, July 1998.

[10] A. Graham, *Kronecker Products and Matrix Calculus with Applications*, Ellis Horwood, Chichester, England, 1981.

[11] J. W. Brewer, "Kronecker products and matrix calculus in system theory," *IEEE Transactions on Circuits and Systems*, vol. CAS-25, no. 9, pp. 772–781, September 1978.

[12] P. Stoica and R. L. Moses, *Introduction to Spectral Analysis*, Prentice Hall, Upper Saddle River, NJ, 1997.

[13] T. Söderström and P. Stoica, *System Identification*, Prentice Hall International, London, UK, 1989.

[14] J. R. Schott, *Matrix Analysis for Statistics*, John Wiley & Sons, Inc., New York, NY, 1997.

[15] E. J. Hannan and B. Wahlberg, "Convergence rates for inverse Toeplitz matrix forms," *Journal of Multivariate Analysis*, vol. 31, pp. 127–135, October 1989.

[16] A. Mitra and P. Stoica, "The asymptotic Cramér-Rao bound for 2-D superimposed exponential signals," *Multidimensional Systems and Signal Processing*, 2000, submitted.

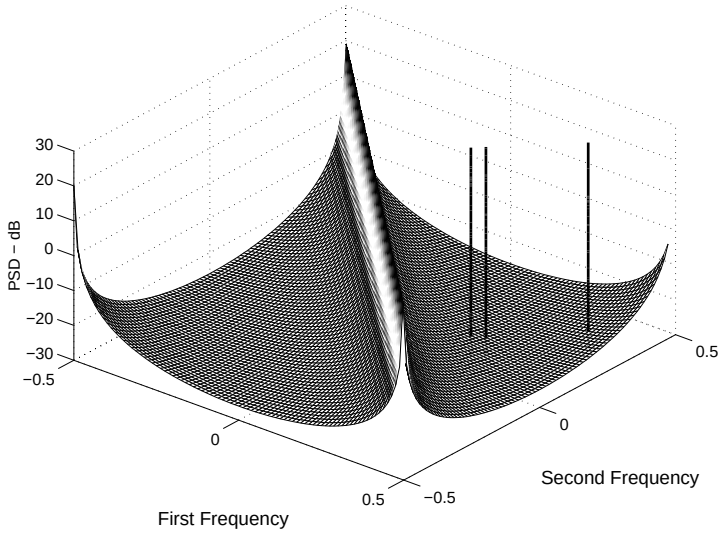
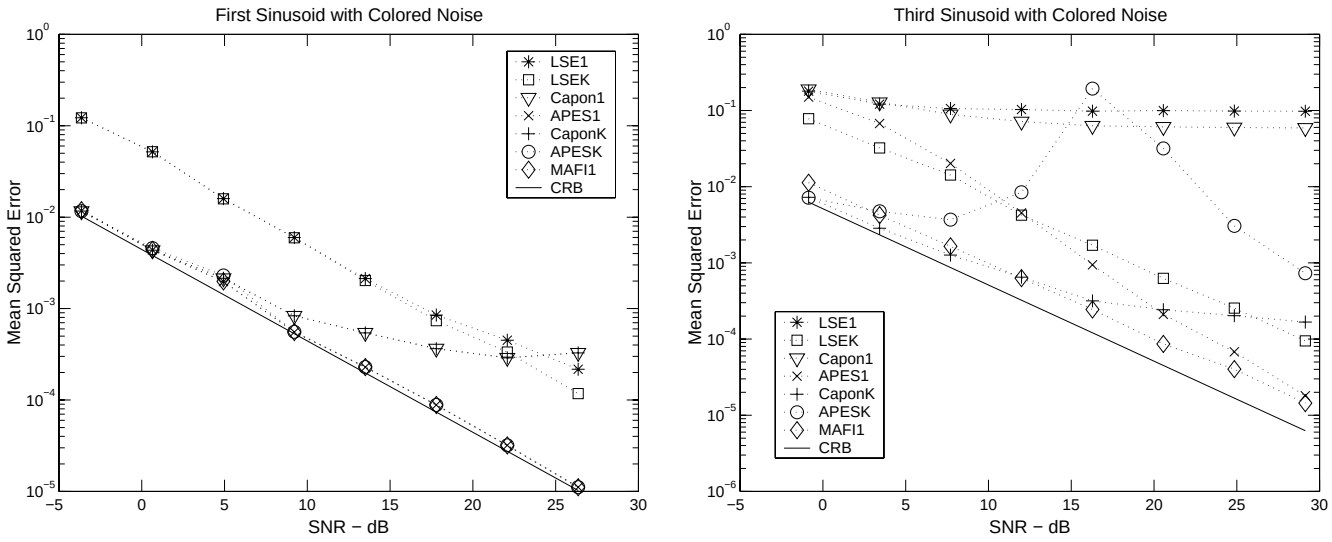
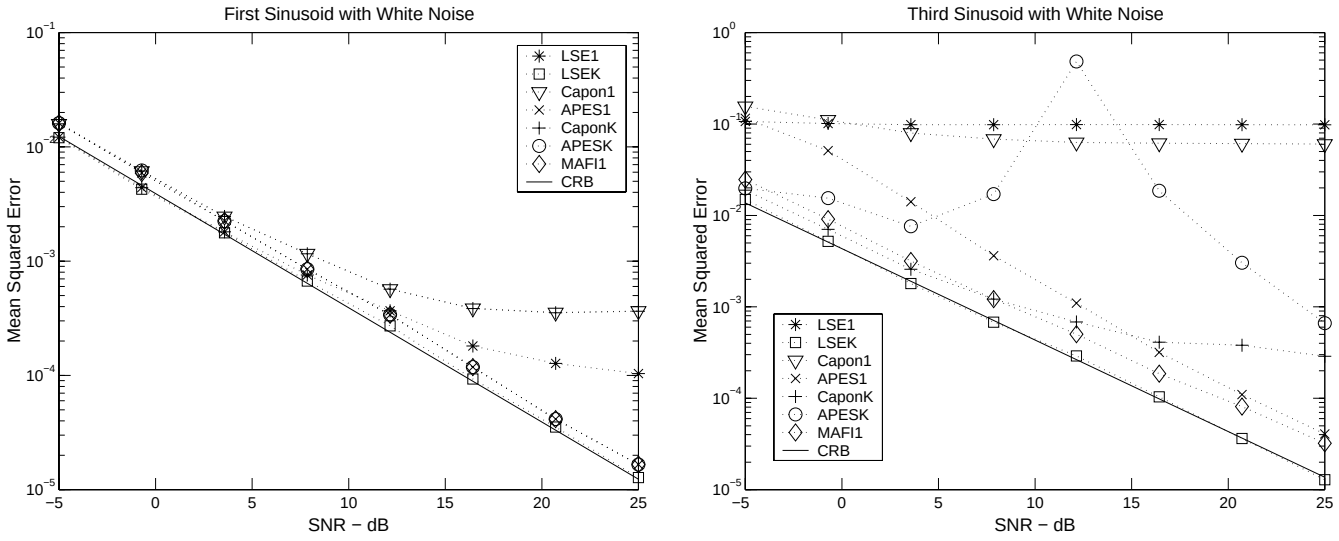


Figure 1: PSD of the test data consisting of three 2-D sinusoids and a 2-D AR noise.



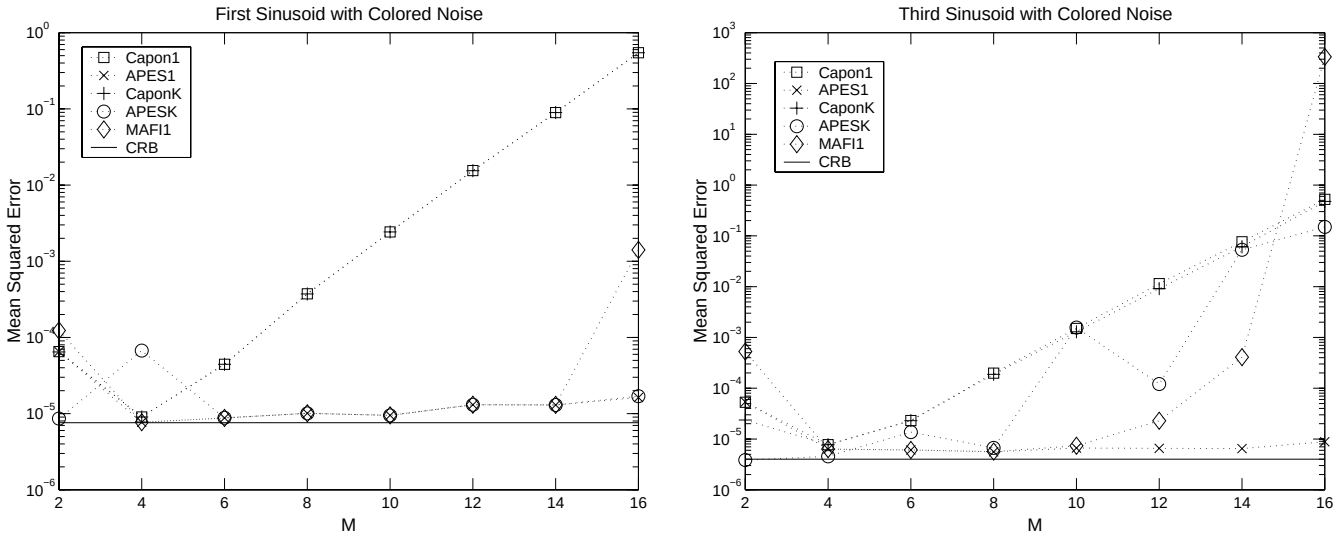
(a) (b)

Figure 2: Empirical MSE and CRB versus SNR with $N = \bar{N} = 16$, $M = \bar{M} = 4$, and colored noise. (a) α_1. (b) α_3.



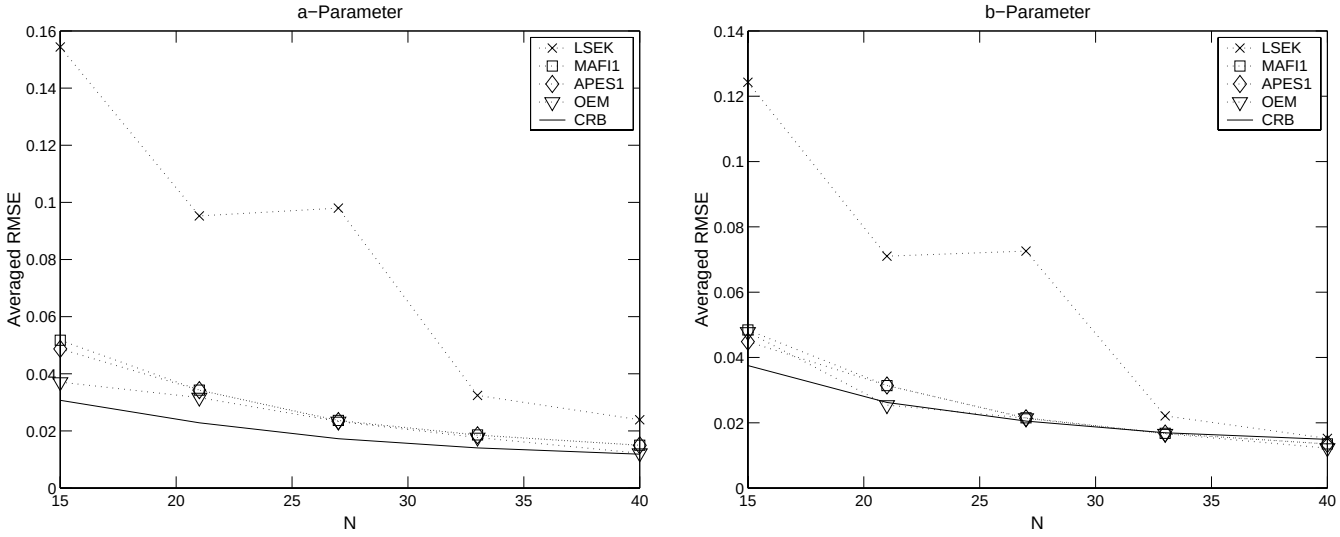
(a) (b)

Figure 3: Empirical MSE and CRB versus SNR with $N = \bar{N} = 16$, $M = \bar{M} = 4$, and white noise. (a) α_1. (b) α_3.



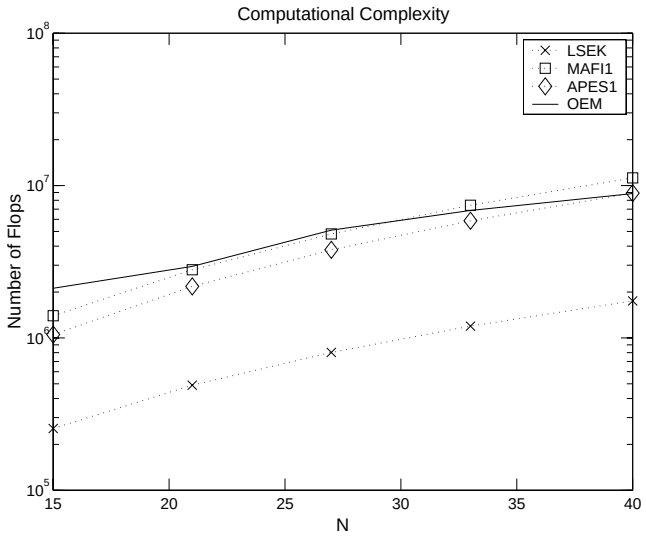
(a) (b)

Figure 4: Empirical MSE and CRB versus $M = \bar{M}$ with $N = \bar{N} = 32$ and colored noise. (a) α_1. (b) α_3.



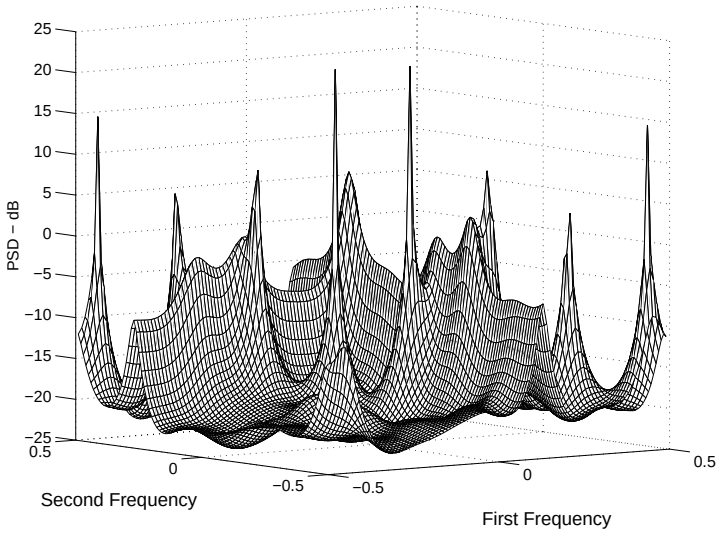
(a)

(b)

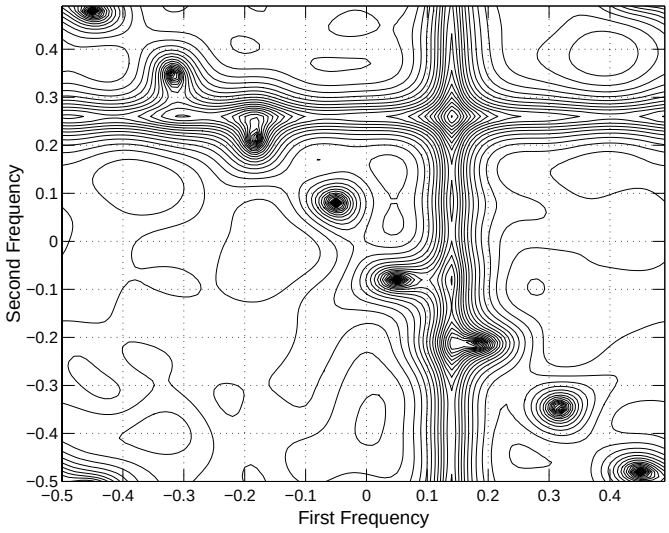


(c)

Figure 5: Averaged RMSE and the number of flops versus $N = \bar{N}$ with $M = \bar{M} = 5$ and white noise. (a) a parameters. (b) b parameters. (c) Number of flops.

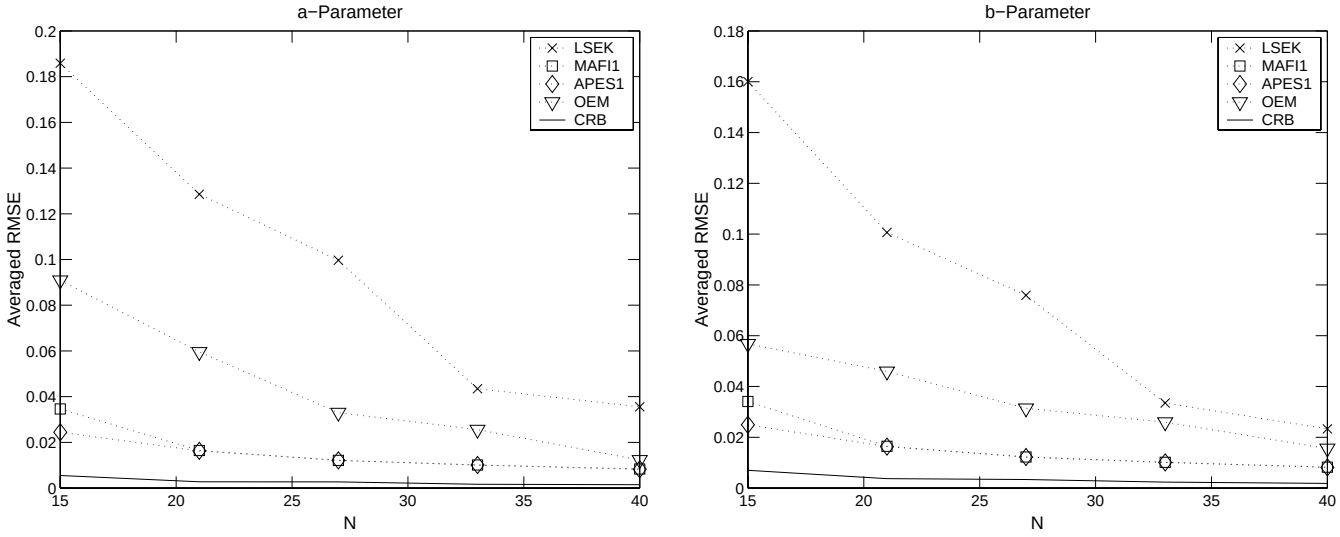


(a)



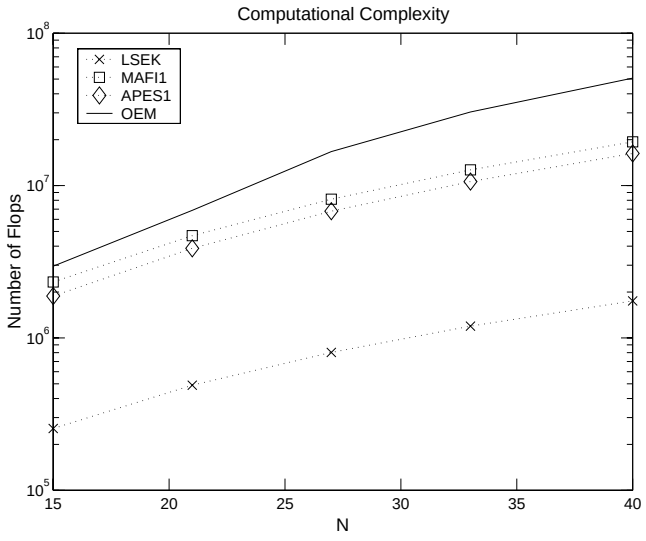
(b)

Figure 6: 2-D PSD estimate and the associated contour plot of the system output with $\sigma^2 = 0.01$, $N = \bar{N} = 25$, and white noise. (a) PSD estimate. (b) Contour.



(a)

(b)



(c)

Figure 7: Averaged RMSE and the number of flops versus $N = \bar{N}$ with $M = \bar{M} = 5$ and colored noise. (a) a parameters. (b) b parameters. (c) Number of flops.