Strong Consistency of the Over- and Underdetermined LSE of 2-D Exponentials in White Noise

Mark Kliger, *Student Member, IEEE,* and Joseph M. Francos, *Senior Member, IEEE*

***Abstract*—We consider the problem of least squares estimation of the parameters of two–dimensional (2-D) exponential signals observed in the presence of an additive noise field, when the assumed number of exponentials is incorrect. We consider both the case where the number of exponential signals is underestimated, and the case where the number of exponential signals is overestimated. In the case where the number of exponential signals is underestimated, we prove the almost sure convergence of the least squares estimates (LSE) to the parameters of the dominant exponentials. In the case where the number of exponential signals is overestimated, the estimated parameter vector obtained by the least squares estimator contains a subvector that converges almost surely to the correct parameters of the exponentials.**

***Index Terms*—Least squares estimation, model-order selection, random fields, strong consistency, two–dimensional (2-D) exponentials, 2-D parameter estimation.**

I. Introduction

In this correspondence, we consider the problem of estimating the parameters of two–dimensional (2-D) exponential signals, observed in the presence of an additive noise field. This problem is, in fact, a special case of the more general problem of estimating the parameters of a 2-D regular and homogeneous random field from a single observed realization of it, Francos [5], Francos *et al.* [3]. This modeling and estimation problem has fundamental theoretical importance, as well as various applications in texture estimation of images (see, e.g., [4] and the references therein) and in wave propagation problems (see, e.g., [18] and the references therein).

The problem of estimating 2-D exponential signals has been intensively investigated in the literature (see, e.g., Priestley [16], Lang and McClellan [13], Kumaresan and Tufts [11], Hua [8], Yang and Hua [20], Chun and Bose [1], Kay and Nekovei [9], Kundu and Gupta [12], Rao *et al.* [17], Francos *et al.* [3], Li and Stoica [14], Li *et al.* [15]). Recently, Rao *et al.* [17] have studied the asymptotic properties of the maximum-likelihood estimator (MLE) of 2-D exponential signals observed in noise. In this framework, the strong consistency of the least squares estimates (LSE) of the parameters of 2-D exponentials observed in the presence of complex white circular Gaussian noise, has been proved. Kundu and Gupta [12] have extended the result of [17] to the case where the observation noise is not necessarily Gaussian. In both papers, as well as in most of the previous studies, it is assumed that the number of exponentials is known. However, this assumption does not always hold in practice.

In this correspondence, we consider the problem of *least squares* estimation of the parameters of 2-D exponential signals observed in the presence of an additive noise field, when the assumed number of exponentials is incorrect. Let P denote the number of exponential signals in the observed field and let k denote their assumed number. In the case where the number of exponential signals is underestimated, i.e., $k < P$, we prove the almost sure convergence of the LSE to the parameters of the k dominant exponentials. In the case where the number of exponential signals is overestimated, i.e., $k > P$, we prove the almost sure convergence of the estimates obtained by the least squares estimator to the parameters of the P exponentials in the observed field. The additional $k - P$ components, assumed to exist, are assigned by the least squares estimator to the dominant components of the periodogram of the noise field.

A solution to the problem addressed here, is an essential component in the error analysis of the least squares (LS) algorithm for estimating 2-D exponentials in noise and in analyzing the performance of the model order selection criterion [10].

Manuscript received January 21, 2004; revised May 13, 2005.

The authors are with the Department of Electrical and Computer Engineering, Ben-Gurion University, Beer-Sheva 84105, Israel (e-mail:kliger@ee.bgu.ac.il; francos@ee.bgu.ac.il).

Communicated by A. Kavčić, Associate Editor for Detection and Estimation.
Digital Object Identifier 10.1109/TIT.2005.853311

II. Notations, Definitions, and Assumptions

Let $\{y(n,m)\}$ be a complex valued field

$$y(n,m) = \sum_{i=1}^{P} a_i^0 e^{j(\omega_i^0 n + \upsilon_i^0 m)} + u(n,m) \tag{1}$$

where $0 \le n \le S-1, 0 \le m \le T-1$ and for each i, a_i^0 is nonzero. Due to physical considerations it is further assumed that for each i, $|a_i^0|$ is bounded.

We make the following assumptions.

Assumption 1: The field $\{u(n,m)\}$ is an independent and identically distributed (i.i.d.) complex-valued zero-mean random field. Let

$$u(n,m) = \Re(u(n,m)) + j\Im(u(n,m))$$

where $u_R(n,m) = \Re(u(n,m))$ and $u_I(n,m) = \Im(u(n,m))$ are the real and imaginary parts of $u(n,m)$, respectively. Both $u_R(n,m)$ and $u_I(n,m)$ are zero mean with finite second-order moment $\frac{\sigma^2}{2}$. The real and imaginary parts are independent.

Assumption 2: The spatial frequencies

$$(\omega_i^0, \upsilon_i^0) \in (0, 2\pi) \times (0, 2\pi), \qquad 1 \le i \le P$$

are pairwise different. In other words, $\omega_i^0 \ne \omega_j^0$ or $\upsilon_i^0 \ne \upsilon_j^0$, when $i \ne j$.

Define the loss function due to the error of the kth-order regression model

$$\mathcal{L}(a_1,\ldots,a_k,\omega_1,\upsilon_1,\ldots,\omega_k,\upsilon_k) = \frac{1}{ST}\sum_{n=0}^{S-1}\sum_{m=0}^{T-1}\left| y(n,m) - \sum_{i=1}^{k} a_i e^{j(n\omega_i + m\upsilon_i)}\right|^2. \tag{2}$$

Let $\{\Psi_i\}$ be a sequence of rectangles such that

$$\Psi_i = \{(n,m) \in \mathcal{Z}^2 \mid 0 \le n \le S_i - 1, 0 \le m \le T_i - 1\}.$$

Definition 1: The sequence of subsets $\{\Psi_i\}$ is said to tend to infinity (we adopt the notation $\Psi_i \to \infty$) as $i \to \infty$ if $\lim_{i\to\infty} \min(S_i, T_i) = \infty$ and $0 < \lim_{i\to\infty}(S_i/T_i) < \infty$. To simplify notations, we shall omit in the following the subscript i. Thus, the notation $\Psi(S,T) \to \infty$ implies that both S and T tend to infinity as functions of i, and at roughly the same rate.

Definition 2: Let Θ_k be a bounded and closed subset of the $3k$-dimensional space $\mathrm{C}^k \times ((0,2\pi) \times (0,2\pi))^k$ where for any vector

$$\theta_k = (a_1,\ldots,a_k,\omega_1,\upsilon_1,\ldots,\omega_k,\upsilon_k) \in \Theta_k$$

the coordinate a_i is nonzero and absolutely bounded for every $1 \le i \le k$ while the pairs (ω_i, υ_i) are pairwise different, so that no two regressors coincide. We shall refer to Θ_k as the *parameter space*.

From the model definition and the above assumptions it is clear that

$$\theta_k^0 = (a_1^0,\ldots,a_k^0,\omega_1^0,\upsilon_1^0,\ldots,\omega_k^0,\upsilon_k^0) \in \Theta_k.$$

A vector $\hat{\theta}_k \in \Theta_k$ that minimizes $\mathcal{L}$ is called the *least square estimate* (LSE). In the case where $k = P$, the LSE is a *strongly consistent* estimator of θ_P^0 (see, e.g., [12] and the references therein). In the following sections, we establish the strong consistency of this LSE when the number of exponentials is underestimated or overestimated.

III. Consistency of the LSE for an Underestimated Model Order

Let k denote the assumed number of observed 2-D exponentials, where $k < P$. For any $\delta > 0$, define the set Δ_δ to be a subset of the parameter space Θ_k such that each vector $\theta_k \in \Delta_\delta$ is different from the vector θ_k^0 by at least δ, at least in one of its coordinates, i.e.,

$$\Delta_\delta = \left[\bigcup_{i=1}^{k} A_{i\delta}\right] \cup \left[\bigcup_{i=1}^{k} W_{i\delta}\right] \cup \left[\bigcup_{i=1}^{k} V_{i\delta}\right] \tag{3}$$

where

$$A_{i\delta} = \left\{\theta_k \in \Theta_k : \left|a_i - a_i^0\right| \geq \delta; \delta > 0\right\} \tag{4}$$

$$W_{i\delta} = \left\{\theta_k \in \Theta_k : \left|\omega_i - \omega_i^0\right| \geq \delta; \delta > 0\right\} \tag{5}$$

$$V_{i\delta} = \left\{\theta_k \in \Theta_k : \left|\upsilon_i - \upsilon_i^0\right| \geq \delta; \delta > 0\right\}. \tag{6}$$

To prove the main result of this section we shall need an additional assumption and the following lemmas.

Assumption 3: For convenience, and without loss of generality, we assume that the exponentials are indexed according to a descending order of their amplitudes, i.e.,

$$\left|a_1^0\right| \geq \left|a_2^0\right| \geq \cdots \left|a_k^0\right| > \left|a_{k+1}^0\right| \cdots \geq \left|a_P^0\right| > 0 \tag{7}$$

where we assume that for a given k, $|a_k^0| > |a_{k+1}^0|$ to avoid trivial ambiguities resulting from the case where the kth dominant component is not unique.

Lemma 1:

$$\liminf_{\Psi(S,T)\to\infty} \inf_{\theta_k \in \Delta_\delta} \left(\mathcal{L}(\theta_k) - \mathcal{L}(\theta_k^0)\right) > 0 \quad \text{a.s.} \tag{8}$$

See Appendix A for the proof.

Lemma 2: Let $\{x_n, n \geq 1\}$ be a sequence of random variables. Then

$$\Pr\{x_n \leq 0 \text{ i.o.}\} \leq \Pr\{\liminf_{n\to\infty} x_n \leq 0\} \tag{9}$$

where the abbreviation i.o. stands for *infinitely often*.

See Appendix B for the proof.

The next theorem establishes the strong consistency of the least squares estimator in the case where the number of the regressors is lower than the actual number of exponentials.

Theorem 1: Let Assumptions 1–3 be satisfied. Then, the k-regressor parameter vector $\hat{\theta}_k = (\hat{a}_1, \ldots, \hat{a}_k, \hat{\omega}_1, \hat{\upsilon}_1, \ldots, \hat{\omega}_k, \hat{\upsilon}_k)$ that minimizes (2) is a *strongly consistent* estimate of

$$\theta_k^0 = (a_1^0, \ldots, a_k^0, \omega_1^0, \upsilon_1^0, \ldots, \omega_k^0, \upsilon_k^0)$$

as $\Psi(S,T) \to \infty$. That is,

$$\hat{\theta}_k \to \theta_k^0, \quad \text{a.s. as } \Psi(S,T) \to \infty. \tag{10}$$

Proof: The proof follows an argument proposed by Wu [19, Lemma 1]. Let $\hat{\theta}_k = (\hat{a}_1, \ldots, \hat{a}_k, \hat{\omega}_1, \hat{\upsilon}_1, \ldots, \hat{\omega}_k, \hat{\upsilon}_k)$ be a parameter vector that minimizes (2). Assume that the proposition $\hat{\theta}_k \to \theta_k^0$ a.s. as $\Psi(S,T) \to \infty$ is not true. Then, there exists some $\delta > 0$, such that [2, Theorem 4.2.2, p. 73]

$$\Pr(\hat{\theta}_k \in \Delta_\delta \text{ i.o.}) > 0. \tag{11}$$

This inequality together with the definition of $\hat{\theta}_k$ as a vector that minimizes $\mathcal{L}$ implies

$$\Pr\left(\inf_{\theta_k \in \Delta_\delta} (\mathcal{L}(\theta_k)) - \mathcal{L}\left(\theta_k^0\right) \leq 0 \text{ i.o.}\right) > 0. \tag{12}$$

Using Lemma 2, we obtain

$$\Pr\left(\liminf_{\Psi(S,T)\to\infty} \inf_{\theta_k\in\Delta_\delta} \left(\mathcal{L}\left(\theta_k\right) - \mathcal{L}\left(\theta_k^0\right)\right) \leq 0\right) \geq \Pr\left(\inf_{\theta_k \in \Delta_\delta} \left(\mathcal{L}(\theta_k) - \mathcal{L}\left(\theta_k^0\right)\right) \leq 0 \text{ i.o.}\right) > 0 \tag{13}$$

which contradicts (8). Hence,

$$\hat{\theta}_k \to \theta_k^0 \quad \text{a.s. as } \Psi(S,T) \to \infty. \tag{14}$$

□

IV. Consistency of the LSE for an Overestimated Model Order

Let k denote the assumed number of observed 2-D exponentials, where $k > P$. Without loss of generality, we can assume that $k = P+1$ (as the proof for $k \geq P+1$ follows immediately by repeating the same arguments). Let the periodogram of the field $\{u(n,m)\}$ be given by

$$I_u(\omega, \upsilon) = \frac{1}{ST}\left|\sum_{n=0}^{S-1}\sum_{m=0}^{T-1} u(n,m) e^{-j(n\omega + m\upsilon)}\right|^2. \tag{15}$$

The parameter spaces Θ_P, Θ_{P+1} are defined as in Definition 2.

To prove the main result of this section we need an additional assumption.

Assumption 4: The real and imaginary components of $u(n,m)$ are such that

$$E[u_R(0,0)^2 \log|u_R(0,0)|] < \infty$$

and

$$E[u_I(0,0)^2 \log|u_I(0,0)|] < \infty.$$

(For example, a white Gaussian noise field satisfies this assumption.)

Theorem 2: Let Assumptions 1, 2, and 4 be satisfied. Then, the parameter vector

$$\hat{\theta}_{P+1} = (\hat{a}_1, \ldots, \hat{a}_P, \hat{a}_{P+1}, \hat{\omega}_1, \hat{\upsilon}_1, \ldots, \hat{\omega}_P, \hat{\upsilon}_P, \hat{\omega}_{P+1}, \hat{\upsilon}_{P+1}) \in \Theta_{P+1}$$

that minimizes (2) with $k = P+1$ regressors as $\Psi(S,T) \to \infty$ is composed of the vector

$$\hat{\theta}_P = (\hat{a}_1, \ldots, \hat{a}_P, \hat{\omega}_1, \hat{\upsilon}_1, \ldots, \hat{\omega}_P, \hat{\upsilon}_P)$$

which is a *strongly consistent* estimate of

$$\theta_P^0 = (a_1^0, \ldots, a_P^0, \omega_1^0, \upsilon_1^0, \ldots, \omega_P^0, \upsilon_P^0)$$

as $\Psi(S,T) \to \infty$; of the pair of spatial frequencies $(\hat{\omega}_{P+1}, \hat{\upsilon}_{P+1})$ that maximizes the periodogram of the observed realization of the field $\{u(n,m)\}$, i.e.,

$$(\hat{\omega}_{P+1}, \hat{\upsilon}_{P+1}) = \arg \max_{(\omega,\upsilon)\in(0,2\pi)^2} I_u(\omega,\upsilon) \tag{16}$$

and of the element $\hat{a}_{P+1}$ that satisfies

$$|\hat{a}_{P+1}|^2 = \frac{1}{ST} I_u(\hat{\omega}_{P+1}, \hat{\upsilon}_{P+1}). \tag{17}$$

Proof: Let

$$\theta_{P+1} = (a_1,\ldots,a_P,a_{P+1},\omega_1,\upsilon_1,\ldots,\omega_P,\upsilon_P,\omega_{P+1},\upsilon_{P+1})$$

be some vector in the parameter space Θ_{P+1}. The LS function with $P+1$ regressors will be denoted $\mathcal{L}_{P+1}$ and the LS function with P regressors will be denoted $\mathcal{L}_P$. We have

$$\begin{aligned}
\mathcal{L}_{P+1}&(\theta_{P+1}) \\
&= \frac{1}{ST}\sum_{n=0}^{S-1}\sum_{m=0}^{T-1}\left|y(n,m) - \sum_{i=1}^{P+1} a_i e^{j(n\omega_i + m\upsilon_i)}\right|^2 \\
&= \frac{1}{ST}\sum_{n=0}^{S-1}\sum_{m=0}^{T-1}\left|y(n,m) - \sum_{i=1}^{P} a_i e^{j(n\omega_i + m\upsilon_i)}\right|^2 \\
&\quad + \frac{1}{ST}\sum_{n=0}^{S-1}\sum_{m=0}^{T-1}\left|a_{P+1} e^{j(n\omega_{P+1} + m\upsilon_{P+1})}\right|^2 \\
&\quad - 2\Re\left\{\frac{1}{ST}\sum_{n=0}^{S-1}\sum_{m=0}^{T-1}\left(y(n,m) - \sum_{i=1}^{P} a_i e^{j(n\omega_i + m\upsilon_i)}\right)\left(a_{P+1} e^{j(n\omega_{P+1} + m\upsilon_{P+1})}\right)^*\right\} \\
&= \mathcal{L}_P(\theta_P) + |a_{P+1}|^2 - 2\Re\left\{\frac{1}{ST}\sum_{n=0}^{S-1}\sum_{m=0}^{T-1} u(n,m) \times \left(a_{P+1} e^{j(n\omega_{P+1} + m\upsilon_{P+1})}\right)^*\right\} \\
&\quad - 2\Re\left\{\frac{1}{ST}\sum_{n=0}^{S-1}\sum_{m=0}^{T-1}\left(\sum_{i=1}^{P} a_i^0 e^{j(n\omega_i^0 + m\upsilon_i^0)} - \sum_{i=1}^{P} a_i e^{j(n\omega_i + m\upsilon_i)}\right)\left(a_{P+1} e^{j(n\omega_{P+1} + m\upsilon_{P+1})}\right)^*\right\} \\
&= H_1(\theta_{P+1}) + H_2(\theta_{P+1}) + H_3(\theta_{P+1})
\end{aligned} \tag{18}$$

where $\theta_P = (a_1,\ldots,a_P,\omega_1,\upsilon_1,\ldots,\omega_P,\upsilon_P) \in \Theta_P$ and

$$\begin{aligned}
H_1(\theta_{P+1}) &= \mathcal{L}_P(a_1,\ldots,a_P,\omega_1,\upsilon_1,\ldots,\omega_P,\upsilon_P) \\
&= \mathcal{L}_P(\theta_P)
\end{aligned} \tag{19}$$

$$H_2(\theta_{P+1}) = |a_{P+1}|^2 - 2\Re\left\{\frac{1}{ST}\sum_{n=0}^{S-1}\sum_{m=0}^{T-1} u(n,m) \times \left(a_{P+1} e^{j(n\omega_{P+1} + m\upsilon_{P+1})}\right)^*\right\} \tag{20}$$

$$H_3(\theta_{P+1}) = -2\Re\left\{\frac{1}{ST}\sum_{n=0}^{S-1}\sum_{m=0}^{T-1}\left(\sum_{i=1}^{P} a_i^0 e^{j(n\omega_i^0 + m\upsilon_i^0)} - \sum_{i=1}^{P} a_i e^{j(n\omega_i + m\upsilon_i)}\right)\left(a_{P+1} e^{j(n\omega_{P+1} + m\upsilon_{P+1})}\right)^*\right\}. \tag{21}$$

Let $\hat{\theta}_P = (\hat{a}_1,\ldots,\hat{a}_P,\hat{\omega}_1,\hat{\upsilon}_1,\ldots,\hat{\omega}_P,\hat{\upsilon}_P)$ be a vector in Θ_P that minimizes $H_1(\theta_{P+1}) = \mathcal{L}_P(\theta_P)$. From [12] (or using Theorem 1 in the previous section)

$$\hat{\theta}_P \to \theta_P^0 \quad \text{a.s. as } \Psi(S,T) \to \infty. \tag{22}$$

The function H_2 is a function of $a_{P+1}, \omega_{P+1}, \upsilon_{P+1}$ only. Evaluating the partial derivatives of H_2 with respect to these variables, it is easy to verify that the extremum points of H_2 are also the extremum points of the periodogram of the realization of the noise field. Moreover, let $a^e, \omega^e, \upsilon^e$ denote an extremum point of H_2. Then at this point

$$H_2(a^e,\omega^e,\upsilon^e) = -\frac{I_u(\omega^e,\upsilon^e)}{ST}. \tag{23}$$

Hence, the minimal value of H_2 is obtained at the coordinates $a_{P+1}, \omega_{P+1}, \upsilon_{P+1}$ where the periodogram of $\{u(n,m)\}$ is maximal. Let $\hat{a}_{P+1}, \hat{\omega}_{P+1}, \hat{\upsilon}_{P+1}$ denote the coordinates that minimize H_2. Then we have

$$\begin{aligned}
(\hat{\omega}_{P+1}, \hat{\upsilon}_{P+1}) &= \arg \min_{(\omega,\upsilon)\in(0,2\pi)^2} H_2(a_{P+1},\omega_{P+1},\upsilon_{P+1}) \\
&= \arg \max_{(\omega,\upsilon)\in(0,2\pi)^2} I_u(\omega,\upsilon)
\end{aligned} \tag{24}$$

and

$$\hat{a}_{P+1} = \frac{1}{ST}\sum_{n=0}^{S-1}\sum_{m=0}^{T-1} u(n,m) e^{-j(n\hat{\omega}_{P+1} + m\hat{\upsilon}_{P+1})} \tag{25}$$

By Assumption 4 [7, Theorem 2.2], we have

$$\sup_{\omega,\upsilon} I_u(\omega,\upsilon) = O(\log ST). \tag{26}$$

Therefore,

$$H_2(\hat{a}_{P+1}, \hat{\omega}_{P+1}, \hat{\upsilon}_{P+1}) = O\left(\frac{\log ST}{ST}\right). \tag{27}$$

Let $\hat{\theta}_{P+1} \in \Theta_{P+1}$ be the vector composed of the elements of the vector $\hat{\theta}_P \in \Theta_P$ and of $\hat{a}_{P+1}, \hat{\omega}_{P+1}, \hat{\upsilon}_{P+1}$, defined above, i.e.,

$$\hat{\theta}_{P+1} = (\hat{a}_1,\ldots,\hat{a}_P,\hat{a}_{P+1},\hat{\omega}_1,\hat{\upsilon}_1,\ldots,\hat{\omega}_P,\hat{\upsilon}_P,\hat{\omega}_{P+1},\hat{\upsilon}_{P+1}).$$

We need to verify that this vector minimizes $\mathcal{L}_{P+1}(\theta_{P+1})$ on Θ_{P+1} as $\Psi(S,T) \to \infty$.

Recall that for $\omega \in (0,2\pi)$

$$\sum_{n=0}^{N-1} e^{j\omega n} = O(1). \tag{28}$$

Hence, as $N \to \infty$

$$\frac{1}{\log N}\sum_{n=0}^{N-1} e^{j\omega n} = o(1) \tag{29}$$

and consequently

$$\frac{1}{N}\sum_{n=0}^{N-1} e^{j\omega n} = o\left(\frac{\log N}{N}\right). \tag{30}$$

Next, we evaluate H_3. Consider the first term in (21). By (30) and unless there exists some $i, 1 \le i \le P$, such that $(\omega_{P+1}, \upsilon_{P+1}) = (\omega_i^0, \upsilon_i^0)$, we have as $\Psi(S,T) \to \infty$

$$\frac{1}{ST}\sum_{n=0}^{S-1}\sum_{m=0}^{T-1}\sum_{i=1}^{P} a_i^0 e^{j(n\omega_i^0 + m\upsilon_i^0)} \times \left(a_{P+1} e^{j(n\omega_{P+1} + m\upsilon_{P+1})}\right)^* = o\left(\frac{\log ST}{ST}\right) \tag{31}$$

for *any* set of values $a_{P+1}, \omega_{P+1}, \upsilon_{P+1}$ may assume.

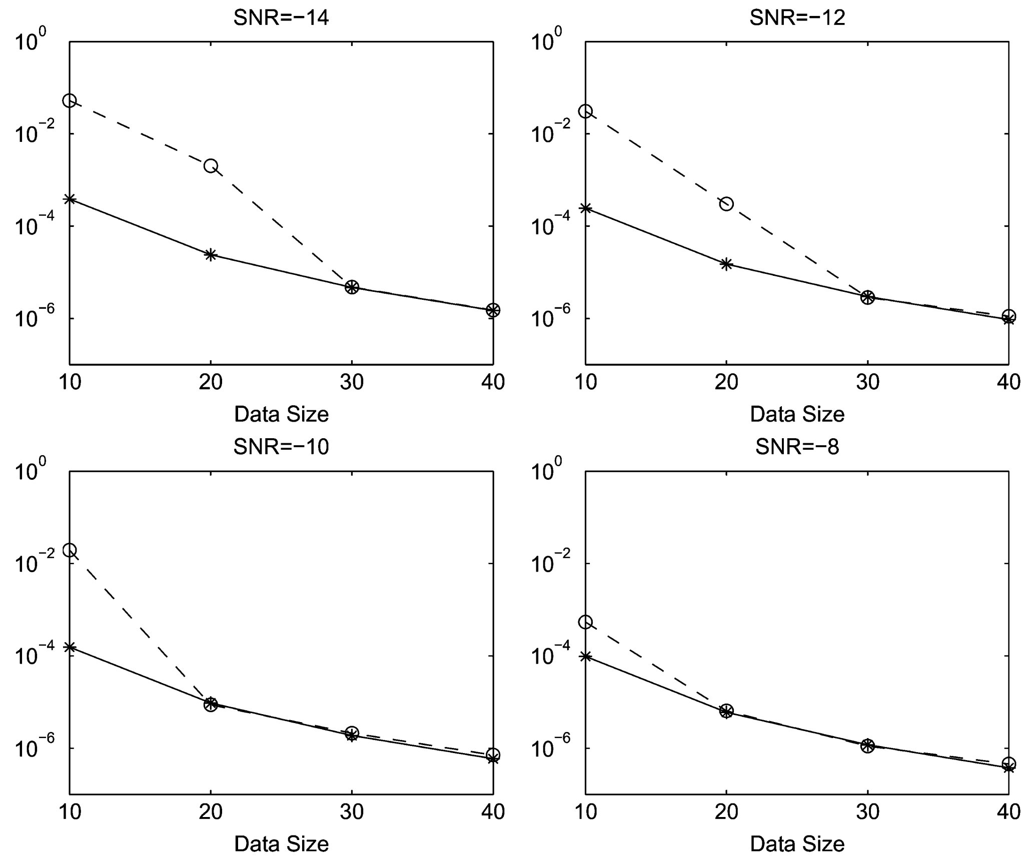

Fig. 1. Error variance (dashed line) in estimating the frequency of the dominant exponential when the model order is underestimated, compared with the corresponding exact CRB (solid line).

Assume now that there exists some i, $1 \leq i \leq P$, such that $(\omega_{P+1}, \upsilon_{P+1}) = (\omega_i^0, \upsilon_i^0)$. Since, by assumption, there are exactly P exponentials, while there are no two different regressors with identical spatial frequencies, it follows that one of the frequency pairs in the estimated vector θ_{P+1}, say (ω_i, υ_i), will have to coincide with a noise peak. As indexing of the components is arbitrary, by interchanging the roles of $(\omega_{P+1}, \upsilon_{P+1})$ and (ω_i, υ_i), and repeating the above argument, we conclude that this term has the same order as in (31). Similarly, for the second term in (21): By (30) and unless there exists some $i, 1 \leq i \leq P$, such that $(\omega_{P+1}, \upsilon_{P+1}) = (\omega_i, \upsilon_i)$, we have as $\Psi(S,T) \to \infty$

$$\frac{1}{ST}\sum_{n=0}^{S-1}\sum_{m=0}^{T-1}\sum_{i=1}^{P} a_i e^{j(n\omega_i + m\upsilon_i)} \times \left(a_{P+1}e^{j(n\omega_{P+1}+m\upsilon_{P+1})}\right)^* = o\left(\frac{\log ST}{ST}\right). \quad (32)$$

However, such i for which $(\omega_{P+1}, \upsilon_{P+1}) = (\omega_i, \upsilon_i)$ cannot exist, as this amounts to reducing the number of regressors from $P+1$ to P, as two of them coincide. Hence, for *any* $\theta_{P+1} \in \Theta_{P+1}$ as $\Psi(S,T) \to \infty$

$$H_3(\theta_{P+1}) = o\left(\frac{\log ST}{ST}\right). \quad (33)$$

On the other hand, the strong consistency (22) of the LSE under the correct model order assumption implies that as $\Psi(S,T) \to \infty$ the minimal value of $\mathcal{L}_P(\theta_P) = \sigma^2$ almost surely, while from (27) we have for the minimal value of H_2 that $H_2(\theta_{P+1}) = O((\log ST)/ST)$. Hence, the value of $H_3(\theta_{P+1})$ at *any* point in Θ_{p+1} is negligible even relative to the values $\mathcal{L}_P(\theta_P)$ and $H_2(\theta_{P+1})$ assume at their respective *minimum* points. Therefore, evaluating (18) as $\Psi(S,T) \to \infty$ we have

$$\begin{aligned}\mathcal{L}_{P+1}(\theta_{P+1}) &= \mathcal{L}_P(\theta_P) + H_2(a_{P+1}, \omega_{P+1}, \upsilon_{P+1}) + H_3(\theta_{P+1}) \\ &= \mathcal{L}_P(\theta_P) + H_2(a_{P+1}, \omega_{P+1}, \upsilon_{P+1}) + o\left(\frac{\log ST}{ST}\right). \quad (34)\end{aligned}$$

Since $\mathcal{L}_P(\theta_P)$ is a function of the parameter vector θ_P and is independent of $a_{P+1}, \omega_{P+1}, \upsilon_{P+1}$, while H_2 is a function of $a_{P+1}, \omega_{P+1}, \upsilon_{P+1}$ and is independent of θ_P, the problem of minimizing $\mathcal{L}_{P+1}(\theta_{P+1})$ becomes *separable* as $\Psi(S,T) \to \infty$. Thus, minimizing (34) is equivalent to separately minimizing $\mathcal{L}_P(\theta_P)$ and $H_2(a_{P+1}, \omega_{P+1}, \upsilon_{P+1})$ as $\Psi(S,T) \to \infty$. Using the foregoing conclusions, the theorem follows. □

V. Numerical Examples

To illustrate the results of the previous sections, we next present some numerical examples. In the following experiments, the error variances in estimating the exponentials frequencies are compared with the corresponding *exact* Cramer–Rao bound (CRB) which is computed for the *correct* model order, [6].

In the first example, the data field is composed of two exponential signals ($P = 2$) observed in noise, i.e.,

$$y(n,m) = a_1^0 e^{j(\omega_1^0 n + \upsilon_1^0 m)} + a_2^0 e^{j(\omega_2^0 n + \upsilon_2^0 m)} + u(n,m).$$

To illustrate the performance of the LSE when the model order is underestimated, the LS estimate is computed assuming the existence of only

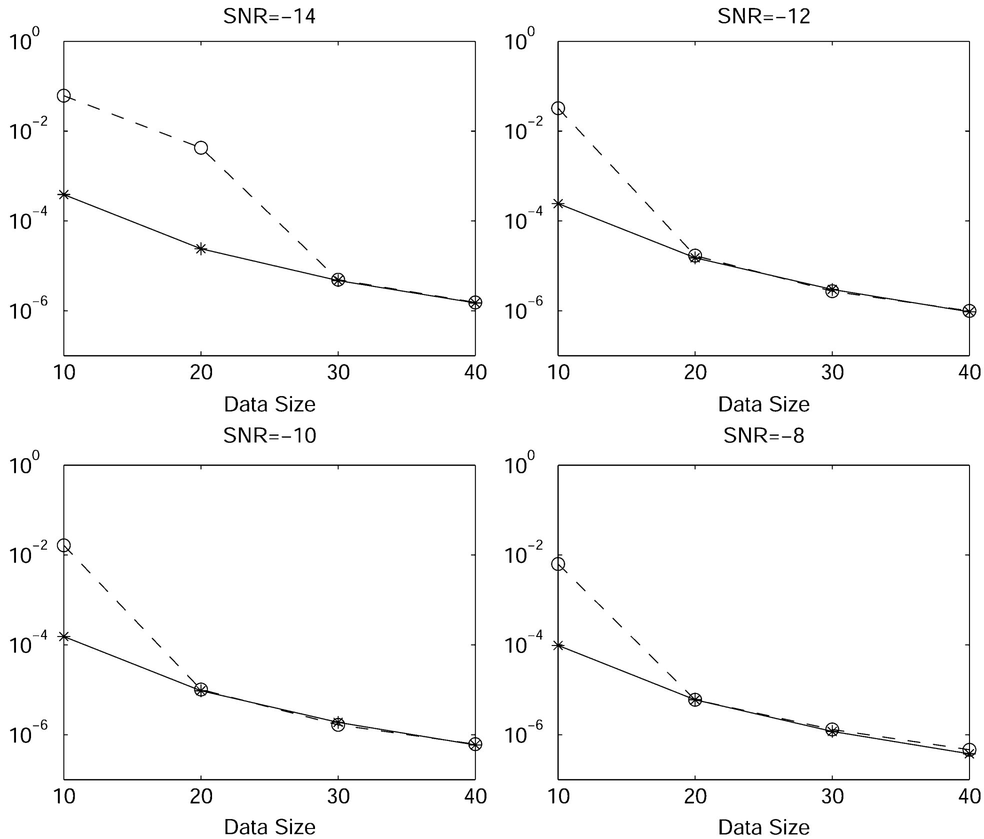

Fig. 2. Error variance (dashed line) in estimating the frequency of an actual exponential when the model order is overestimated, compared with the corresponding exact CRB (solid line).

a single exponential in the observed field $(k=1)$. The signal-to-noise ratio (SNR) for the first component is defined by

$$\mathrm{SNR} = 10\log_{10}\frac{\left|a_1^0\right|^2}{\sigma^2}\ \mathrm{dB}. \tag{35}$$

The noise field $\{u(n,m)\}$ is a complex-valued zero-mean white Gaussian noise field with variance σ^2 (the real and imaginary parts are independent real valued Gaussian white noise fields, each with variance $(\sigma^2/2)$) which is chosen to yield the desired SNR. In this experiment, the data field dimensions range from 10×10 to 40×40. The experiment was performed for four different SNR values. For each data size and each SNR value, 200 Monte Carlo experiments were performed. The frequencies of the exponential signals are $\omega_1^0 = \upsilon_1^0 = 2\pi 0.13$ and $\omega_2^0 = \upsilon_2^0 = 2\pi 0.31$. The amplitudes are chosen such that $|a_1^0|^2 = 3$ and $|a_2^0|^2 = 1$. Initialization of the nonlinear LS minimization procedure, in search for the unknown frequencies, is implemented by evaluating the maximum of the periodogram of the zero-padded observation. The variance of the estimation error of the ω_1^0 frequency parameter of the dominant exponential is shown on Fig. 1 for various data sizes.

In the second example, the data field is composed of a single exponential signal $(P=1)$ observed in noise, i.e.,

$$y(n,m) = a_1^0 e^{j(\omega_1^0 n + \upsilon_1^0 m)} + u(n,m)$$

with parameters identical to those of the first example. The assumed number of exponential signals is two $(k=2)$. The variance of the estimation error of the ω_1^0 frequency parameter of the exponential is shown in Fig. 2 for various data sizes.

Both experiments indicate that despite the incorrect assumptions on the model order, even for modest data dimension and relatively low SNR values the error variance of the LSE achieves the corresponding exact CRB—computed for the *correct* model order. Moreover, note that for the first exponential, the foregoing theoretical results imply the almost sure convergence of the estimated parameters to the correct ones both in the case where the model order is underestimated, as well as in the case where the model order is overestimated. Indeed, it is clear that the LSE performance illustrated in Figs. 1 and 2 is almost identical, despite the different assumptions regarding the model order in each case.

Nevertheless, we note that the above examples provide only an indirect intuition for the results proven earlier, rather than a direct illustration. It is obvious that it is impossible to illustrate almost sure convergence by finite sample examples with finite precision computations. The essence of almost sure convergence is that it tells us that as data size tends to infinity the set of events in which incorrect estimates are obtained is of zero probability. This implies that if one conducts an infinite number of experiments, at most a finite number of estimates will deviate from the correct parameters. However, since the parameters of the exponentials are bounded, the almost sure convergence implies convergence in the mean-square sense. Hence, based on the results proven in the previous sections, and the normality of the noise field in this example, one can expect to observe the convergence of the estimation error variances to the CRB, as has been demonstrated by the above examples.

VI. Discussion and Conclusion

We have considered the problem of least squares estimation of the parameters of 2-D exponential signals observed in the presence of an additive noise field, when the assumed number of exponentials is incorrect. In the case where the number of exponential signals is underestimated, we have proved the almost sure convergence of the LSE to the parameters of the dominant exponentials. This result can be intuitively explained using the basic principles of least squares estimation: Since the least squares estimate is the set of model parameters that minimizes the ℓ_2 norm of the error between the observations and the assumed model, it follows that in the case where the model order is underestimated the minimum error norm is achieved when the k most dominant exponentials are correctly estimated. Similarly, in the case where the number of exponential signals is overestimated, the estimated parameter vector obtained by the least squares estimator contains a $3P$-dimensional subvector that converges almost surely to the correct parameters of the exponentials, while the remaining $k-P$ components assumed to exist, are assigned to the $k-P$ most dominant spectral peaks of the noise power to further minimize the norm of the estimation error.

Appendix A

Lemma 1:

$$\liminf_{\Psi(S,T)\to\infty} \inf_{\theta_k\in\Delta_\delta} \left(\mathcal{L}(\theta_k)-\mathcal{L}\left(\theta_k^0\right)\right)>0 \quad \text{a.s.} \tag{36}$$

Proof: In the following, we first show that on Δ_δ the sequence $\mathcal{L}(\theta_k)-\mathcal{L}(\theta_k^0)$ (indexed in S,T) is uniformly lower-bounded by a strictly positive constant as $\Psi(S,T)\to\infty$. Since the sequence elements are uniformly lower-bounded by a strictly positive constant, the sequence of infimums, $\inf_{\theta_k\in\Delta_\delta}(\mathcal{L}(\theta_k)-\mathcal{L}(\theta_k^0))$ is uniformly lower-bounded by the same strictly positive constant as $\Psi(S,T)\to\infty$, and hence,

$$\liminf_{\Psi(S,T)\to\infty} \inf_{\theta_k\in\Delta_\delta} (\mathcal{L}(\theta_k)-\mathcal{L}(\theta_k^0)).$$

Thus, we first prove that the sequence $\mathcal{L}(\theta_k)-\mathcal{L}(\theta_k^0)$ is uniformly lower-bounded away from zero on Δ_δ as $\Psi(S,T)\to\infty$

$$\begin{aligned}
\mathcal{L}(\theta_k)-\mathcal{L}\left(\theta_k^0\right) &= \frac{1}{ST}\sum_{n=0}^{S-1}\sum_{m=0}^{T-1}\left|y(n,m)-\sum_{i=1}^{k}a_i e^{j(n\omega_i+m\upsilon_i)}\right|^2 \\
&\quad -\frac{1}{ST}\sum_{n=0}^{S-1}\sum_{m=0}^{T-1}\left|y(n,m)-\sum_{i=1}^{k}a_i^0 e^{j(n\omega_i^0+m\upsilon_i^0)}\right|^2 \\
&= \frac{1}{ST}\sum_{n=0}^{S-1}\sum_{m=0}^{T-1}\left|\sum_{i=1}^{P}a_i^0 e^{j(n\omega_i^0+m\upsilon_i^0)}+u(n,m)-\sum_{i=1}^{k}a_i e^{j(n\omega_i+m\upsilon_i)}\right|^2 \\
&\quad -\frac{1}{ST}\sum_{n=0}^{S-1}\sum_{m=0}^{T-1}\left|\sum_{i=k+1}^{P}a_i^0 e^{j(n\omega_i^0+m\upsilon_i^0)}+u(n,m)\right|^2 \\
&= \frac{1}{ST}\sum_{n=0}^{S-1}\sum_{m=0}^{T-1}\left|\sum_{i=1}^{k}a_i^0 e^{j(n\omega_i^0+m\upsilon_i^0)}-\sum_{i=1}^{k}a_i e^{j(n\omega_i+m\upsilon_i)}\right|^2 \\
&\quad +2\Re\left\{\frac{1}{ST}\sum_{n=0}^{S-1}\sum_{m=0}^{T-1}\left(\sum_{i=k+1}^{P}a_i^0 e^{j(n\omega_i^0+m\upsilon_i^0)}\right)\times\left(\sum_{i=1}^{k}a_i^0 e^{j(n\omega_i^0+m\upsilon_i^0)}-\sum_{i=1}^{k}a_i e^{j(n\omega_i+m\upsilon_i)}\right)^*\right\} \\
&\quad +2\Re\left\{\frac{1}{ST}\sum_{n=0}^{S-1}\sum_{m=0}^{T-1}u(n,m)\left(\sum_{i=1}^{k}a_i^0 e^{j(n\omega_i^0+m\upsilon_i^0)}-\sum_{i=1}^{k}a_i e^{j(n\omega_i+m\upsilon_i)}\right)^*\right\} \\
&= I_1+I_2+I_3.
\end{aligned} \tag{37}$$

Thus, to check the asymptotic behavior of the left-hand side of (37), we have to evaluate $\lim_{\Psi(S,T)\to\infty}(I_1+I_2+I_3)$ for all vectors $\theta_k\in\Delta_\delta$

$$\begin{aligned}
\lim_{\Psi(S,T)\to\infty} I_1 &= \lim_{\Psi(S,T)\to\infty}\frac{1}{ST}\sum_{n=0}^{S-1}\sum_{m=0}^{T-1}\left|\sum_{i=1}^{k}a_i^0 e^{j(n\omega_i^0+m\upsilon_i^0)}\right|^2 \\
&\quad + \lim_{\Psi(S,T)\to\infty}\frac{1}{ST}\sum_{n=0}^{S-1}\sum_{m=0}^{T-1}\left|\sum_{i=1}^{k}a_i e^{j(n\omega_i+m\upsilon_i)}\right|^2 \\
&\quad - \lim_{\Psi(S,T)\to\infty}\left[2\Re\left\{\frac{1}{ST}\sum_{n=0}^{S-1}\sum_{m=0}^{T-1}\sum_{i=1}^{k}\sum_{l=1}^{k}a_i^0 a_l^* \times e^{j(n[\omega_i^0-\omega_l]+m[\upsilon_i^0-\upsilon_l])}\right\}\right] \\
&= M_1+M_2+M_3.
\end{aligned} \tag{38}$$

Recall that for $0<|a|<\infty$

$$\lim_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N-1}ae^{j\omega n}=0 \tag{39}$$

uniformly on any closed interval in $(0,2\pi)$. Hence, due to Assumption 2 and (39), we have

$$\begin{aligned}
M_1 &= \lim_{\Psi(S,T)\to\infty}\frac{1}{ST}\sum_{n=0}^{S-1}\sum_{m=0}^{T-1}\left|\sum_{i=1}^{k}a_i^0 e^{j(n\omega_i^0+m\upsilon_i^0)}\right|^2 \\
&= \sum_{i=1}^{k}\left|a_i^0\right|^2
\end{aligned} \tag{40}$$

independently of θ_k.

Also,

$$\begin{aligned}
M_2 &= \lim_{\Psi(S,T)\to\infty}\frac{1}{ST}\sum_{n=0}^{S-1}\sum_{m=0}^{T-1}\left|\sum_{i=1}^{k}a_i e^{j(n\omega_i+m\upsilon_i)}\right|^2 \\
&= \sum_{i=1}^{k}|a_i|^2+\lim_{\Psi(S,T)\to\infty}\frac{1}{ST} \times\sum_{n=0}^{S-1}\sum_{m=0}^{T-1}\sum_{\substack{i=1\\i\neq l}}^{k}\sum_{l=1}^{k}a_i a_l^* e^{j(n[\omega_i-\omega_l]+m[\upsilon_i-\upsilon_l])}.
\end{aligned} \tag{41}$$

Since the pairs (ω_i,υ_i) are pairwise different, then on any closed interval in $(0,2\pi)$ the sequence of partial sums

$$\frac{1}{ST}\sum_{n=0}^{S-1}\sum_{m=0}^{T-1}\sum_{\substack{i=1\\i\neq l}}^{k}\sum_{l=1}^{k}a_i a_l^* e^{j(n[\omega_i-\omega_l]+m[\upsilon_i-\upsilon_l])}$$

converges uniformly to zero as $\Psi(S,T)\to\infty$.

Hence,

$$M_2=\sum_{i=1}^{k}|a_i|^2 \tag{42}$$

uniformly on Δ_δ as $\Psi(S,T)\to\infty$.

Leaving M_3 unchanged, we obtain

$$\begin{aligned}
\lim_{\Psi(S,T)\to\infty} I_1 &= \sum_{i=1}^{k}\left(\left|a_i^0\right|^2+|a_i|^2\right) \\
&\quad -\lim_{\Psi(S,T)\to\infty}2\Re\left\{\frac{1}{ST}\sum_{n=0}^{S-1}\sum_{m=0}^{T-1}\sum_{i=1}^{k}\sum_{l=1}^{k}a_i^0 a_l^* \times e^{j(n[\omega_i^0-\omega_l]+m[\upsilon_i^0-\upsilon_l])}\right\}
\end{aligned} \tag{43}$$

uniformly on Δ_δ.

Using similar considerations to those employed in the evaluation of (40) we obtain

$$\lim_{\Psi(S,T)\to\infty} I_2 = \lim_{\Psi(S,T)\to\infty} 2\Re\left\{\frac{1}{ST}\sum_{n=0}^{S-1}\sum_{m=0}^{T-1}\left(\sum_{i=k+1}^{P} a_i^0 e^{j(n\omega_i^0+m\upsilon_i^0)}\right)\right.$$
$$\left.\times\left(\sum_{i=1}^{k} a_i^0 e^{j(n\omega_i^0+m\upsilon_i^0)} - \sum_{i=1}^{k} a_i e^{j(n\omega_i+m\upsilon_i)}\right)^*\right\}$$
$$= -\lim_{\Psi(S,T)\to\infty} 2\Re\left\{\frac{1}{ST}\sum_{n=0}^{S-1}\sum_{m=0}^{T-1}\sum_{i=k+1}^{P}\sum_{l=1}^{k} a_i^0 a_l^*\right.$$
$$\left.\times e^{j(n[\omega_i^0-\omega_l]+m[\upsilon_i^0-\upsilon_l])}\right\}. \quad (44)$$

By a straightforward extension of [12, Lemma 1] we have

$$\sup_{\theta_k\in\Delta_\delta}\left|\frac{1}{ST}\sum_{n=0}^{S-1}\sum_{m=0}^{T-1} u(n,m)\right.$$
$$\left.\times\left(\sum_{i=1}^{k} a_i^0 e^{j(n\omega_i^0+m\upsilon_i^0)} - \sum_{i=1}^{k} a_i e^{j(n\omega_i+m\upsilon_i)}\right)^*\right| \to 0,$$
$$\text{a.s., as } \Psi(S,T)\to\infty. \quad (45)$$

Hence, $I_3 \to 0$ almost surely as $\Psi(S,T)\to\infty$ uniformly on Δ_δ. Using (43)–(45) we conclude that almost surely

$$\lim_{\Psi(S,T)\to\infty}\left(\mathcal{L}(\theta_k) - \mathcal{L}\left(\theta_k^0\right)\right)$$
$$= \sum_{i=1}^{k}\left(\left|a_i^0\right|^2 + |a_i|^2\right)$$
$$- \lim_{\Psi(S,T)\to\infty} 2\Re\left\{\frac{1}{ST}\sum_{n=0}^{S-1}\sum_{m=0}^{T-1}\sum_{i=1}^{P}\sum_{l=1}^{k} a_i^0 a_l^*\right.$$
$$\left.\times e^{j(n[\omega_i^0-\omega_l]+m[\upsilon_i^0-\upsilon_l])}\right\} \quad (46)$$

uniformly on Δ_δ.

To complete the evaluation of the uniform lower bound, we consider all the vectors $\theta_k \in \Delta_\delta$. Let us initially consider the subset $A_{q\delta} \subset \Delta_\delta$ for some $q, 1 \le q \le k$. Thus, the coordinate a_q of each vector in this subset is different from the corresponding coordinate a_q^0 by at least $\delta > 0$. Consider first the case where all the other elements of the vector $\theta_k \in A_{q\delta}$ are identical to the corresponding elements of θ_k^0. Since by this assumption $\omega_l = \omega_l^0, \upsilon_l = \upsilon_l^0$ for $1 \le l \le k$, and $a_l = a_l^0$, for $1 \le l \le k, l \ne q$, on this set we have

$$\lim_{\Psi(S,T)\to\infty}\left(\mathcal{L}(\theta_k) - \mathcal{L}\left(\theta_k^0\right)\right)$$
$$= \left|a_q - a_q^0\right|^2 - \lim_{\Psi(S,T)\to\infty} 2\Re\left\{\frac{1}{ST}\sum_{n=0}^{S-1}\sum_{m=0}^{T-1}\sum_{\substack{i=1\\ i\ne l}}^{P}\sum_{l=1}^{k} a_i^0 a_l^*\right.$$
$$\left.\times\ e^{j(n[\omega_i^0-\omega_l^0]+m[\upsilon_i^0-\upsilon_l^0])}\right\}$$
$$= \left|a_q^0 - a_q\right|^2 \ge \delta^2 > 0 \quad \text{a.s.} \quad (47)$$

uniformly in a_q, where the second equality is due to Assumption 2 and by following the same arguments employed to obtain (42).

Assume next that $\theta_k \in A_{q\delta}$ (i.e., the coordinate a_q is different from the corresponding coordinate a_q^0 by at least δ) and that in addition, there exists an element a_t of θ_k, such that $1 \le t \le k, t \ne q$ and $|a_t - a_t^0| \ge \lambda, \lambda > 0$ while all the other elements of the vector θ_k are identical to the corresponding elements of θ_k^0. Following a similar derivation to the one in (47) we conclude that

$$\lim_{\Psi(S,T)\to\infty}\left(\mathcal{L}(\theta_k) - \mathcal{L}\left(\theta_k^0\right)\right) = \left|a_q^0 - a_q\right|^2 + \left|a_t^0 - a_t\right|^2$$
$$\ge \delta^2 + \lambda^2 > \delta^2 \quad \text{a.s.} \quad (48)$$

uniformly in a_q and a_t.

Finally, consider the case where $\theta_k \in A_{q\delta}$ while there exists an element ω_l of θ_k, such that $|\omega_l - \omega_l^0| \ge \eta, \eta > 0$ and all the other elements of the vector θ_k are identical to the corresponding elements of θ_k^0. Following a similar derivation to the one in (47), we conclude that

$$\lim_{\Psi(S,T)\to\infty}\left(\mathcal{L}(\theta_k) - \mathcal{L}\left(\theta_k^0\right)\right)$$
$$= \begin{cases}\left|a_q^0 - a_q\right|^2 + 2\left|a_l^0\right|^2, & l \ne q\\ \left|a_q^0\right|^2 + |a_q|^2, & l = q\end{cases} > \delta^2 \quad \text{a.s.} \quad (49)$$

uniformly in a_q and ω_l.

From the above analysis it is clear that almost surely $\lim_{\Psi(S,T)\to\infty}(\mathcal{L}(\theta_k) - \mathcal{L}(\theta_k^0))$ is lower-bounded by δ^2 uniformly on $A_{q\delta}$.

Following similar reasoning, the next subset we consider is $W_{q\delta} \cup V_{q\delta}$. We first consider a subset of this set

$$\Lambda = \left\{\theta_k \in W_{q\delta} \cup V_{q\delta} : \exists p, k+1 \le p \le P, (\omega_q, \upsilon_q) = \left(\omega_p^0, \upsilon_p^0\right)\right\}$$
$$\subset W_{q\delta} \cup V_{q\delta}. \quad (50)$$

This subset includes vectors in Θ_k, such that their coordinate pair (ω_q, υ_q) is different from the corresponding pair of θ_k^0 and equal to some pair $(\omega_p^0, \upsilon_p^0)$ where $p \ge k+1$. As above, the minimum is obtained when all the other elements of θ_k are identical to the corresponding elements of θ_k^0. Hence, uniformly on Λ, we have

$$\lim_{\Psi(S,T)\to\infty}\left(\mathcal{L}(\theta_k) - \mathcal{L}(\theta_k^0)\right) \ge |a_q^0|^2 + |a_q|^2 - 2\Re\{a_p^0 a_q^*\}$$
$$= \left|a_q^0\right|^2 - \left|a_p^0\right|^2 + |a_q - a_p^0|^2$$
$$\ge \left|a_q^0\right|^2 - \left|a_p^0\right|^2 = \epsilon_\Lambda > 0 \text{ a.s.} \quad (51)$$

where the last inequality is due to Assumption 3.

On the complementary set

$$\Lambda^c = (W_{q\delta} \cup V_{q\delta})\backslash\Lambda$$
$$= \left\{\theta_k \in W_{q\delta} \cup V_{q\delta} : (\omega_q, \upsilon_q)\right.$$
$$\left.\ne \left(\omega_p^0, \upsilon_p^0\right),\ \forall p, k+1 \le p \le P\right\} \quad (52)$$

we have

$$\lim_{\Psi(S,T)\to\infty}\left(\mathcal{L}(\theta_k) - \mathcal{L}\left(\theta_k^0\right)\right) \ge \left|a_q^0\right|^2 + |a_q|^2 \ge \left|a_q^0\right|^2$$
$$= \epsilon_{\Lambda^c} > 0 \text{ a.s.} \quad (53)$$

Let $\epsilon_q = \min(\delta^2, \epsilon_\Lambda, \epsilon_{\Lambda^c})$. Collecting (47), (51), and (53) together we conclude that almost surely the sequence $\mathcal{L}(\theta_k) - \mathcal{L}(\theta_k^0)$ is lower-bounded by $\epsilon_q > 0$ uniformly on $A_{q\delta} \cup W_{q\delta} \cup V_{q\delta}$ as $\Psi(S,T)\to\infty$.

By repeating the same arguments for every $q, 1 \le q \le k$, and by letting $\epsilon = \min(\epsilon_1, \ldots, \epsilon_k)$, we conclude that almost surely . the sequence $\mathcal{L}(\theta_k) - \mathcal{L}(\theta_k^0)$ (indexed in S, T) is lower-bounded by $\epsilon > 0$ uniformly on Δ_δ as $\Psi(S,T)\to\infty$.

Hence, it follows that sequence $\inf_{\theta_k \in \Delta_\delta}(\mathcal{L}(\theta_k) - \mathcal{L}(\theta_k^0))$ (indexed in S, T) is also asymptotically lower-bounded by $\epsilon > 0$, i.e.,

$$\inf_{\theta_k \in \Delta_\delta} \left(\mathcal{L}(\theta_k) - \mathcal{L}\left(\theta_k^0\right)\right) \geq \epsilon \quad \text{a.s.} \tag{54}$$

as $\Psi(S,T) \to \infty$.

Hence, by the definition of $\liminf$

$$\liminf_{\Psi(S,T)\to\infty} \inf_{\theta_k \in \Delta_\delta} \left(\mathcal{L}(\theta_k) - \mathcal{L}\left(\theta_k^0\right)\right) \geq \epsilon > 0 \quad \text{a.s.} \tag{55}$$

□

Appendix B

Lemma 2: Let $\{x_n, n \geq 1\}$ be a sequence of random variables. Then

$$\Pr\{x_n \leq 0 \text{ i.o.}\} \leq \Pr\{\liminf_{n\to\infty} x_n \leq 0\}. \tag{56}$$

Proof: Let (Ω, Σ, p) be some probability space. Let $\{x_n(\omega), n \geq 1\}$ be a sequence of random variables. Let $\{A_n \in \Sigma, n \geq 1\}$ be a sequence of subsets of Ω, such that $A_n = \{\omega \in \Omega : x_n(\omega) \leq 0\}$. Define

$$A_n^m = \bigcup_{n=m}^{\infty} \{\omega : x_n \leq 0\}. \tag{57}$$

Then

$$A_n^m \subseteq \{\omega : \inf_{n\geq m} x_n \leq 0\}. \tag{58}$$

Hence,

$$\bigcap_{m=1}^{\infty} A_n^m \subseteq \bigcap_{m=1}^{\infty} \{\omega : \inf_{n\geq m} x_n \leq 0\}. \tag{59}$$

Consider the right-hand side of (59), and let $y_m(\omega) = \inf_{n\geq m} x_n$. Since for all $\omega \in \bigcap_{m=1}^{\infty}\{\omega : \inf_{n\geq m} x_n \leq 0\}, y_m(\omega) \leq 0$ for all m, then by definition $\sup_m y_m(\omega) \leq 0$ as well. On the other hand, if $\sup_m y_m(\omega) \leq 0$, then for all $m, y_m(\omega) \leq 0$. Hence, we have the following set equality:

$$\bigcap_{m=1}^{\infty} \{\omega : \inf_{n\geq m} x_n \leq 0\} = \{\omega : \sup_m \inf_{n\geq m} x_n \leq 0\}. \tag{60}$$

Rewriting (59) we have

$$\begin{aligned}\bigcap_{m=1}^{\infty} \bigcup_{n=m}^{\infty} A_n &\subseteq \{\omega : \sup_m \inf_{n\geq m} x_n \leq 0\} \\ &= \{\omega : \liminf_{n\to\infty} x_n(\omega) \leq 0\}\end{aligned} \tag{61}$$

where the equality on the right-hand side of (61) follows from the definition of $\liminf_{n\to\infty}(\cdot)$ of a sequence x_n. Also, by definition,

$$\bigcap_{m=1}^{\infty} \bigcup_{n=m}^{\infty} A_n = \limsup_{n\to\infty} A_n.$$

Hence (see, e.g., [2, p. 72])

$$\begin{aligned}\limsup_{n\to\infty} A_n &= \{\omega : x_n(\omega) \leq 0 \text{ i.o.}\} \\ &\subseteq \{\omega : \liminf_{n\to\infty} x_n(\omega) \leq 0\}.\end{aligned} \tag{62}$$

Due to the monotonicity of the probability measure, the lemma follows. □

References

[1] J. Chun and N. K. Bose, "Parameter estimation via signal-selectivity of signal-subspace (PESS) and its application to 2-D wavenumber estimation," *Digital Signal Process.*, vol. 5, pp. 58–76, 1995.

[2] K. L. Chung, *A Course in Probability Theory*. New York: Academic, 1974.

[3] J. M. Francos, A. Narashimhan, and J. W. Woods, "Maximum-likelihood parameter estimation of discrete homogeneous random fields with mixed spectral distribution," *IEEE Trans. Signal Process.*, vol. 44, no. 5, pp. 1242–1255, May 1996.

[4] ——, "Maximum-likelihood estimation of textures using a wold decomposition model," *IEEE Trans. Image Process.*, vol. 4, no. 12, pp. 1655–1666, Dec. 1995.

[5] J. M. Francos, "Bounds on the accuracy of estimating the parameters of discrete homogeneous random fields with mixed spectral distributions," *IEEE Trans. Inf. Theory*, vol. 43, no. 3, pp. 908–922, May 1997.

[6] ——, "Cramer-Rao bound on the estimation accuracy of complex valued homogeneous Gaussian random fields," *IEEE Trans. Signal Process.*, vol. 50, no. 3, pp. 710–724, Mar. 2002.

[7] S. He, "Uniform convergency for weighted periodogram of stationary linear random fields," *Chin. Ann. Math.*, vol. 16B, pp. 331–340, 1995.

[8] Y. Hua, "Estimating two-dimensional frequencies by matrix enhancement and matrix pencil," *IEEE Trans. Signal Process.*, vol. 40, no. 9, pp. 2267–2280, Sep. 1992.

[9] S. Kay and R. Nekovei, "An efficient two-dimensional frequency estimator," *IEEE Trans. Signal Process.*, vol. 38, no. 10, pp. 1807–1809, Oct. 1990.

[10] M. Kliger and J. M. Francos, "MAP model order selection rule for 2-D sinusoids in white noise," *IEEE Trans. Signal Process.*, vol. 53, no. 7, pp. 2563–2575, Jul. 2005.

[11] R. Kumaresan and D. W. Tufts, "A two-dimensional technique for frequency-wavenumber estimation," *Proc. IEEE*, vol. 69, no. 11, pp. 1515–1517, Nov. 1981.

[12] D. Kundu and R. D. Gupta, "Asymptotic properties of the least squares estimates of a two dimensional model," *Metrika*, vol. 48, pp. 83–97, 1998.

[13] S. W. Lang and J. H. McClellan, "The extension of Pisarenko's method to multiple dimensions," in *Proc. Int. Conf. Acoustics, Speech, Signal Processing*, Paris, France, May 1982, pp. 125–128.

[14] J. Li and P. Stoica, "Efficient mixed-spectrum estimation with applications to target feature extraction," *IEEE Trans. Signal Process.*, vol. 44, no. 2, pp. 281–295, Feb. 1996.

[15] J. Li, P. Stoica, and D. Zheng, "An efficient algorithm for tow-dimensional frequency estimation," *Multidim. Systems and Signal Process.*, vol. 7, pp. 151–178, 1996.

[16] M. B. Priestley, *Spectral Analysis and Time Series*. New York: Academic, 1981.

[17] C. R. Rao, L. Zhao, and B. Zhou, "Maximum likelihood estimation of 2-D superimposed exponential signals," *IEEE Trans. Signal Process.*, vol. 42, no. 7, pp. 1795–1802, Jul. 1994.

[18] J. Ward, "Space-Time Adaptive Processing for Airborne Radar," Lincoln Lab., MIT, Lexington, MA, Tech. Rep. 1015, 1994.

[19] C. F. J. Wu, "Asymptotic theory of nonlinear least squares estimation," *Ann. Statist.*, vol. 9, pp. 510–513, 1981.

[20] H. Yang and Y. Hua, "Statistical decomposition of an eigendecomposition based method for 2-D frequency estimation," *Automatica*, vol. 30, pp. 157–168, 1994.