STRONGLY CONSISTENT MODEL ORDER SELECTION FOR ESTIMATING 2-D SINUSOIDS IN COLORED NOISE

Mark Kliger and Joseph M. Francos

Electrical and Computer Engineering Department
Ben-Gurion University
Beer Sheva 84105, Israel

ABSTRACT

The problem of jointly estimating the number as well as the parameters of two-dimensional sinusoidal signals, observed in the presence of an additive colored noise field is considered. We begin by establishing the strong consistency of the non-linear least squares estimator of the parameters of two-dimensional sinusoids, when the number of sinusoidal signals assumed in the field is under-estimated, or over-estimated. Based on these results, we prove the strong consistency of a new family of model order selection rules.

1. INTRODUCTION

We consider the problem of jointly estimating the number as well as the parameters of two-dimensional sinusoidal signals, observed in the presence of an additive colored noise field.

This problem is, in fact, a special case of a much more general problem, [8]: From the 2-D Wold-like decomposition we have that any 2-D regular and homogeneous discrete random field (analogous of the 1-D wide-sense stationary process) can be represented as a sum of two mutually orthogonal components: a purely- indeterministic field and a deterministic one. In this paper we consider the special case where the deterministic component consists of a finite (unknown) number of sinusoidal components, while the purely-indeterministic component is an infinite order non-symmetrical half plane, (or a quarter-plane), moving average (MA) field (colored noise field). This modeling and estimation problem has fundamental theoretical importance, as well as various applications in texture estimation of images (see, *e.g.* [7] and the references therein) and in space-time adaptive processing of airborne radar data (see, *e.g.* [26] and the references therein).

Many algorithms have been devised to estimate the parameters of two-dimensional sinusoids observed in the presence of an additive white noise field and only a small fraction of the derived methods has been extended to the case where the noise field is colored (see, [6], [11], [14], [16], [23], and the references therein). Moreover, most of these algorithms assume the number of sinusoids is *a-priori* known. However this assumption only rarely holds in practice.

In the past several decades the problem of model order selection for 1-D signals has received considerable attention. In general, model order selection rules are based (directly or indirectly) on three popular criteria: Akaike Information Criterion (AIC), [2], the Minimum Description Length (MDL), [21] and the Bayesian Information Criterion (BIC) [22]. All these criteria have a common form composed of two terms: a data term and a penalty term. The data term monotonically decreases as the model order increases. The data term is usually taken to be the negative log-likelihood for an assumed model order, or the variance of the residual component of the least-square regression for an assumed model order. The penalty term is a function (usually linear or log-linear) of the model order and the size of the a data sample. For example, AIC penalty is a linear function of the model order only, while the MDL/BIC penalties are linear functions of the model order and log-linear functions of the size of the data sample. The penalties of MDL and BIC are identical.

In [24] and [25] Zhao et. al. proposed the Efficient Detection Criterion (EDC) for detecting the number of signals observed in white or colored noise. In contrast to the fixed penalties of AIC/MDL/BIC model order selection rules, the penalty term of EDC is not fixed, but rather a family of penalties. The strong consistency of EDC has been proven for the case where the penalty term increase slower than the size of data, but faster than loglog of size of data. For example, MDL/BIC penalty which increases with a rate of log of the size of the data is a member of EDC penalty family.

Due to its theoretical and practical importance in many problems of statistics and signal processing, the question of how to determine the number of 1-D sinusoids observed in the presence of white or colored noises has been extensively investigated (see [5], [15], [17], [19], and the references therein). Quinn, [19], has proved that in the case of 1-D sinusoids observed in white noise AIC/MDL/BIC type model order selection rules lead to consistent order selection only if the penalty function increases with a rate proportional to the log of the size of data and the proportionality constant has a crucial role in the consistency of estimator [19].

The problem of model order selection for multidimensional fields in general, and multidimensional harmonic fields in particular, has received much less attention. Usually one of the standard penalties (MDL/BIC penalties are among the most popular) is applied to solve the model order selection problem for 2-D sinusoids in noise (see, *e.g.* [18]) or other penalties which were derived for the 1-D case are adopted for the 2-D case (see, *e.g.* [17]).

In [12], following ideas of [19], we proved the strong consistency of a large family of model order selection rules *specifically designed* for the case of 2-D sinusoids observed in white Gaussian noise. In the present paper we derive a strongly consistent model order selection rule, for jointly estimating the number of sinusoidal components and their parameters in the presence of *colored* noise. This derivation extends the results of [12] to the case where the additive noise is colored, modeled by an infinite order non-symmetrical half-plane or quarter-plane moving average representation. Moreover, in the case considered in this paper, the noise field is not necessarily Gaussian.

The proposed criterion has the standard form of a data term and a penalty term, where the data term is the variance of the residual of the *least squares estimator* evaluated for the assumed model order (the loss function). It is well known that the non-linear least square estimator of the parameters of 2-D sinusoids in noise is strongly consistent, [14]. However, this result was proven only for a case when the number of sinusoids is *a-priori* known and correct. Since similarly to AIC/MDL/BIC framework, we evaluate the data term for *any* assumed model order, including incorrect ones, we should first address the meaning of consistency of least squares estimation of the parameters of 2-D sinusoidal signals when the assumed number of sinusoids is *incorrect*.

Let P denote the true number of 2-D sinusoidal signals in the observed field and let k denote their assumed number by the least squares estimator of the model parameters. In the case where the number of sinusoidal signals is under-estimated, *i.e.*, $k < P$, we prove in the following the almost sure convergence of the least squares estimates to the parameters of the k dominant sinusoids. In the case where the number of sinusoidal signals is over-estimated, *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 sinusoids 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. These results extend our previous results on the consistency of the least squares estimator of complex exponentials observed in the presence of an additive white noise field [13].

The penalty term of the proposed model order selection rule is proportional to the logarithm of the size of the data sample. Similarly to [19] and [12], the coefficient of proportion has a crucial role in the consistency of estimator. We will prove the strong consistency of the new model order selection criterion and will show how different assumptions regarding the noise field affect the penalty term of the criterion. The proposed criterion completely generalized the previous results [12], and provides a strongly consistent estimator of the number as well as of the parameters of the sinusoidal components.

2. NOTATIONS, DEFINITIONS AND ASSUMPTIONS

We begin by formulating the general framework. Let $\{y(n,m)\}$ be a real valued field,

$$y(n,m) = \sum_{i=1}^{P} \rho_i^0 \cos(\omega_i^0 n + \upsilon_i^0 m + \varphi_i^0) + w(n,m), \quad (1)$$

where $0 \le n \le N-1$, $0 \le m \le M-1$ and for each i, ρ_i^0 is non-zero. Due to physical considerations it is further assumed that for each i, amplitude $|\rho_i^0|$ is bounded .

The noise field $\{w(n,m)\}$ represents the purely-indeterministic component of Wold decomposition and assumed to be an infinite order non-symmetrical half plane moving average (MA) field.

Recall that the *non-symmetrical half-plan total-order* is defined by

$$(i,j) \succeq (s,t) \text{ iff}$$
$$(i,j) \in \{(k,l)|k=s, l \ge t\} \cup \{(k,l)|k>s, -\infty \le l \le \infty\} \quad (2)$$

Let D be an *infinite* order non-symmetrical half-plane support, defined by

$$D = \{(i,j) \in \mathbb{Z}^2 : i=0, 0 \le j \le \infty\} \cup \{(i,j) \in \mathbb{Z}^2 : 0 < i \le \infty, -\infty \le j \le \infty\}. \quad (3)$$

Hence the notations $(r,s) \in D$ and $(r,s) \succeq (0,0)$ are equivalent.

We assume that $\{w(n,m)\}$ is an infinite order non-symmetrical half-plane MA noise field, *i.e.*,

$$w(n,m) = \sum_{(r,s)\in D} a(r,s) u(n-r, m-s), \quad (4)$$

such that the following assumptions are satisfied:

Assumption 1: The field $\{u(n,m)\}$ is an i.i.d. real valued zero-mean random field with finite variance σ^2, such that $E[|u(n,m)|^\alpha] < \infty$ for some $\alpha > 3$.

Assumption 2: The sequence $a(i,j)$ is an absolutely summable deterministic sequence, *i.e.*,

$$\sum_{(r,s)\in D} |a(r,s)| < \infty. \quad (5)$$

Let $f_w(\omega,\upsilon)$ denote the spectral density function of the noise field $\{w(n,m)\}$. Hence,

$$f_w(\omega,\upsilon) = \sigma^2 \left| \sum_{(r,s)\in D} a(r,s) e^{j(\omega r + \upsilon s)} \right|^2. \quad (6)$$

Assumption 3: The spatial frequencies $(\omega_i^0, \upsilon_i^0) \in (0,2\pi) \times (0,2\pi)$, $1 \le i \le P$ are pairwise different. In other words, $\omega_i^0 \ne \omega_j^0$ or $\upsilon_i^0 \ne \upsilon_j^0$, when $i \ne j$.

Let $\{\Psi_i\}$ be a sequence of rectangles such that $\Psi_i = \{(n,m) \in \mathbb{Z}^2 \mid 0 \le n \le N_i - 1, 0 \le m \le M_i - 1\}$.

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

$$\lim_{i\to\infty} \min(N_i, M_i) = \infty,$$

and

$$0 < \lim_{i\to\infty} (N_i/M_i) < \infty.$$

To simplify notations, we shall omit in the following the subscript i. Thus, the notation $\Psi(N,M) \to \infty$ implies that both N and M 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 $4k$ dimensional space $\mathbb{R}^k \times ((0,2\pi)\times(0,2\pi))^k \times [0,2\pi)^k$ where for any vector $\theta_k = (\rho_1, \omega_1, \upsilon_1, \varphi_1, \ldots, \rho_k, \omega_k, \upsilon_k, \varphi_k) \in \Theta_k$ the coordinate ρ_i is non-zero and 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 (1) and the above assumptions it is clear that

$$\theta_k^0 = (\rho_1^0, \omega_1^0, \upsilon_1^0, \varphi_1^0, \ldots, \rho_k^0, \omega_k^0, \upsilon_k^0, \varphi_k^0) \in \Theta_k.$$

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

$$\mathcal{L}_k(\theta_k) = \frac{1}{NM} \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} \left(y(n,m) - \sum_{i=1}^{k} \rho_i^0 \cos(\omega_i^0 n + \upsilon_i^0 m + \varphi_i^0) \right)^2. \quad (7)$$

A vector $\hat{\theta_k} \in \Theta_k$ that minimizes $\mathcal{L}_k(\theta_k)$ is called the *Least Squares Estimate* (LSE). In the case where $k = P$, the LSE is a *strongly consistent* estimator of θ_P^0 (see, *e.g.*, [14] and the references therein).

3. STRONG CONSISTENCY OF THE OVER- AND UNDER-DETERMINED LSE

As mentioned in the Introduction, it is well known that the least squares estimator of the parameters of 2-D sinusoids observed in the presence of colored additive noise field is strongly consistent (see [14]). However, this result relies on the assumption that the correct number of sinusoids is a-priori known. In this section we consider the asymptotic behavior of the LSE when the assumed number of sinusoids is incorrect.

The first theorem establishes the strong consistency of the least squares estimator in the case where the number of the sinusoidal regressors is lower than the actual number of sinusoids. The second theorem establishes the strong consistency of the least squares estimator in the case where the number of the regressors is higher than the actual number of sinusoids. These theorems extend the results proved in [13] for the case where the additive noise field is white and complex-valued.

Let k denote the assumed number of observed 2-D sinusoids, where $k < P$, *i.e.* the number of regressors is lower than the actual number of sinusoids.

In order to establish the next theorem we shall need an additional assumption:

Assumption 4: For convenience, and without loss of generality, we assume that the sinusoids are indexed according to a descending order of their amplitudes, *i.e.*,

$$\rho_1^0 \geq \rho_2^0 \geq \ldots \rho_k^0 > \rho_{k+1}^0 \ldots \geq \rho_P^0 > 0 \;, \tag{8}$$

where we assume that for a given k, $\rho_k^0 > \rho_{k+1}^0$ to avoid trivial ambiguities resulting from the case where the k-th dominant component is not unique.

Theorem 1. *Let Assumptions 1-4 be satisfied. Let $k < P$. Then, the k-regressor parameter vector* $\hat{\theta}_k = (\hat{\rho}_1, \hat{\omega}_1, \hat{\upsilon}_1, \hat{\varphi}_1, \ldots, \hat{\rho}_k, \hat{\omega}_k, \hat{\upsilon}_k, \hat{\varphi}_k)$ *that minimizes (7) is a strongly consistent estimator of* $\theta_k^0 = (\rho_1^0, \omega_1^0, \upsilon_1^0, \varphi_1^0, \ldots, \rho_k^0, \omega_k^0, \upsilon_k^0, \varphi_k^0)$ *as* $\Psi(N, M) \to \infty$. *That is,*

$$\hat{\theta}_k \to \theta_k^0 \;\; a.s. \;\; as \;\; \Psi(N, M) \to \infty. \tag{9}$$

Proof: See [1].

Theorem 1 implies that even in the case where the sinusoidal signals are observed in the presence of additive colored noise, and the number of sinusoidal signals is under-estimated, the least squares estimates converge to the parameters of the dominant sinusoids. 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 (*i.e.* the variance of the residual component), it follows that in the case where the model order is under-estimated the minimum error norm is achieved when the k most dominant sinusoids are correctly estimated. In other words, the variance of the residual component will be minimized if we will remove the k most dominant sinusoids from the data.

Remark: Actually, Theorem 1 remains valid even under less restrictive assumptions regarding the noise field $\{w(n, m)\}$. If the field $\{u(n, m)\}$ is an i.i.d. real valued zero-mean random field with finite variance σ^2, and the sequence $a(i, j)$ is a square summable deterministic sequence, *i.e.*, $\sum_{(r,s)\in D} a^2(r, s) < \infty$, Theorem 1 holds.

Next, we consider the case where the number of the regressors is larger than the actual number of sinusoids. Let k denote the assumed number of observed 2-D sinusoids, 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). The parameter spaces Θ_P, Θ_{P+1} are defined as in Definition 2. Let the *periodogram* (scaled by a factor of 2) of the field $\{w(n, m)\}$ be given by

$$I_w(\omega, \upsilon) = \frac{2}{NM} \left| \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} w(n, m) e^{-j(n\omega + m\upsilon)} \right|^2 . \tag{10}$$

Let $(\omega_{per}, \upsilon_{per})$ denote the pair of spatial frequencies that maximizes the periodogram of the observed realization of $\{w(n, m)\}$, *i.e.*,

$$(\omega_{per}, \upsilon_{per}) = \underset{(\omega,\upsilon)\in(0,2\pi)^2}{\arg\max} \; I_w(\omega, \upsilon). \tag{11}$$

Also let

$$\rho_{per}^2 = \frac{2}{NM} I_w(\omega_{per}, \upsilon_{per}) \;, \tag{12}$$

denote the squared amplitude of the periodogram at its maximum point. φ_{per} denotes the phase at this point.

Theorem 2. *Let Assumptions 1-4 be satisfied. Then, the parameter vector* $\hat{\theta}_{P+1} = (\hat{\rho}_1, \hat{\omega}_1, \hat{\upsilon}_1, \hat{\varphi}_1, \ldots, \hat{\rho}_{P+1}, \hat{\omega}_{P+1}, \hat{\upsilon}_{P+1}, \hat{\varphi}_{P+1}) \in \Theta_{P+1}$ *that minimizes (7) with $k = P + 1$ regressors is a strongly consistent estimator of* $(\rho_1^0, \omega_1^0, \upsilon_1^0, \varphi_1^0, \ldots, \rho_P^0, \omega_P^0, \upsilon_P^0, \varphi_P^0, \rho_{per}, \omega_{per}, \upsilon_{per}, \varphi_{per})$ *as* $\Psi(N, M) \to \infty$. *That is:*

$$\hat{\theta}_{P+1} \to (\theta_P^0, \rho_{per}, \omega_{per}, \upsilon_{per}, \varphi_{per}) \;\; a.s. \;\; as \;\; \Psi(N, M) \to \infty \tag{13}$$

Proof: See [1].

Thus, in the case where the number of sinusoidal signals is over-estimated, the estimated parameter vector obtained by the least squares estimator contains a $4P$-dimensional sub-vector that converges almost surely to the correct parameters of the sinusoidal components, 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.

4. STRONG CONSISTENCY OF A FAMILY OF MODEL ORDER SELECTION RULES

In this section, using the theorems derived in the previous section, we establish the strong consistency of a new model order selection rule.

It is assumed that there are Q competing models, where Q is finite, $Q > P$, and that each competing model $k \in Z_Q = \{0, 1, 2, \ldots, Q - 1\}$ is equiprobable. Following the MDL/BIC framework, define the statistic

$$\chi_\xi(k) = NM \log \mathcal{L}_k(\hat{\theta}_k) + \xi k \log NM, \tag{14}$$

where ξ is some finite constant to be specified later, and $\mathcal{L}_k(\hat{\theta}_k)$ is the minimal value of the error variance of the least squares estimator.

The number of 2-D sinusoids is estimated by minimizing $\chi_\xi(k)$ over $k \in Z_Q$, *i.e.*,

$$\hat{P} = \underset{k\in Z_Q}{\arg\min} \left\{ \chi_\xi(k) \right\}. \tag{15}$$

Let

$$\mathcal{A} := \frac{\left(\sum_{(r,s)\in D} |a(r,s)|\right)^2}{\sum_{(r,s)\in D} a^2(r,s)}. \tag{16}$$

The objective of the next theorem is to prove the asymptotic consistency of the model order selection procedure in (15).

Theorem 3. *Let Assumptions 1-4 be satisfied. Let $\hat{P}$ be given by (15) with $\xi > 14\mathcal{A}$. Then as $\Psi(N,M) \to \infty$*

$$\hat{P} \to P \ \ a.s. \tag{17}$$

Proof. For $k \le P$,

$$\begin{aligned} &\chi_\xi(k-1) - \chi_\xi(k) \\ &= NM \log \mathcal{L}_{k-1}(\hat{\theta}_{k-1}) + \xi(k-1)\log NM \\ &- NM \log \mathcal{L}_k(\hat{\theta}_k) - \xi k \log NM \\ &= NM \log\left(\frac{\mathcal{L}_{k-1}(\hat{\theta}_{k-1})}{\mathcal{L}_k(\hat{\theta}_k)}\right) - \xi \log NM. \end{aligned} \tag{18}$$

From Theorem 1 as $\Psi(N,M) \to \infty$

$$\hat{\theta}_k \to \theta_k^0 \ \text{ a.s.}, \tag{19}$$

and

$$\hat{\theta}_{k-1} \to \theta_{k-1}^0 \ \text{ a.s.} \tag{20}$$

From the definition of $\mathcal{L}_k(\hat{\theta}_k)$, and (19)

$$\begin{aligned} &\mathcal{L}_k(\hat{\theta}_k) = \frac{1}{NM}\sum_{n=0}^{N-1}\sum_{m=0}^{M-1}\left(y(n,m) - \sum_{i=1}^{k}\hat{\rho}_i\cos(\hat{\omega}_i n + \hat{\upsilon}_i m + \hat{\varphi}_i)\right)^2 \\ &= \frac{1}{NM}\sum_{n=0}^{N-1}\sum_{m=0}^{M-1}\left(\sum_{i=1}^{P}\rho_i^0\cos(\omega_i^0 n + \upsilon_i^0 m + \varphi_i^0) + w(n,m)\right. \\ &\left. - \sum_{i=1}^{k}\hat{\rho}_i\cos(\hat{\omega}_i n + \hat{\upsilon}_i m + \hat{\varphi}_i)\right)^2 \xrightarrow[\Psi(N,M)\to\infty]{} \\ &\frac{1}{NM}\sum_{n=0}^{N-1}\sum_{m=0}^{M-1}\left(\sum_{i=k+1}^{P}\rho_i^0\cos(\omega_i^0 n + \upsilon_i^0 m + \varphi_i^0) + w(n,m)\right)^2. \end{aligned} \tag{21}$$

From Appendix C, in [1] we have that as $\Psi(N,M) \to \infty$

$$\sup_{\omega,\upsilon}\left|\frac{1}{NM}\sum_{n=0}^{N-1}\sum_{m=0}^{M-1} w(n,m)\cos(\omega n + \upsilon m)\right| \to 0 \ \text{ a.s.} \tag{22}$$

Recall also that for $\omega \in (0,2\pi)$ and $\varphi \in [0,2\pi)$

$$\sum_{n=0}^{N-1}\cos(\omega n + \varphi) = \frac{\sin\left([N-\frac{1}{2}]\omega + \varphi\right) + \sin\left(\frac{\omega}{2} - \varphi\right)}{2\sin\left(\frac{\omega}{2}\right)} = O(1). \tag{23}$$

Hence, from Assumption 3, (22), (23), and the Strong Law of Large Numbers, we conclude that as $\Psi(N,M) \to \infty$

$$\mathcal{L}_k(\hat{\theta}_k) \to \sigma^2\sum_{(r,s)\in D} a^2(r,s) + \sum_{i=k+1}^{P}\frac{(\rho_i^0)^2}{2} \ \text{ a.s.} \tag{24}$$

and similarly

$$\mathcal{L}_{k-1}(\hat{\theta}_{k-1}) \to \sigma^2\sum_{(r,s)\in D} a^2(r,s) + \sum_{i=k}^{P}\frac{(\rho_i^0)^2}{2} \ \text{ a.s.} \tag{25}$$

Since $\frac{\log NM}{NM}$ tends to zero, as $\Psi(N,M) \to \infty$, then as $\Psi(N,M) \to \infty$

$$\begin{aligned} &(NM)^{-1}(\chi_\xi(k-1) - \chi_\xi(k)) \to \\ &\log\left(1 + \frac{(\rho_k^0)^2}{2\sigma^2\sum_{(r,s)\in D} a^2(r,s) + \sum_{i=k+1}^{P}(\rho_i^0)^2}\right) \text{a.s.} \end{aligned} \tag{26}$$

Since $\log\left(1 + \frac{(\rho_k^0)^2}{2\sigma^2\sum_{(r,s)\in D} a^2(r,s) + \sum_{i=k+1}^{P}(\rho_i^0)^2}\right)$ is strictly positive, then $\chi_\xi(k-1) > \chi_\xi(k)$. Hence, for $k \le P$, the function $\chi_\xi(k)$ is **monotonically decreasing** with k.

We next consider the case where $k = P + l$ for any integer $l \ge 1$.

Employing Theorem 2 and by repeating the arguments made for $l = 1$ for the case of $l > 1$, it is not difficult to show that a.s. as $\Psi(N,M) \to \infty$ (see the proof of Theorem 2 for the derivation)

$$\mathcal{L}_{P+l}(\hat{\theta}_{P+l}) = \mathcal{L}_P(\hat{\theta}_P) - \frac{U_l}{NM} + o\left(\frac{\log NM}{NM}\right), \tag{27}$$

where

$$U_l = \sum_{i=1}^{l} I_w(\omega_i, \upsilon_i), \tag{28}$$

is the sum of the l largest elements of the periodogram of the noise field $\{w(s,t)\}$. Clearly

$$U_l \le l \sup_{\omega,\upsilon} I_u(\omega,\upsilon). \tag{29}$$

From [14] (or using Theorem 1 in the previous section),

$$\hat{\theta}_P \to \theta_P^0 \ \ a.s. \ \text{ as } \ \Psi(N,M) \to \infty. \tag{30}$$

Hence, the strong consistency (30) of the LSE under the correct model order assumption implies that as $\Psi(N,M) \to \infty$

$$\mathcal{L}_P(\hat{\theta}_P) \to \sigma^2\sum_{(r,s)\in D} a^2(r,s) \ \text{ a.s.} \tag{31}$$

Thus, almost surely as $\Psi(N,M) \to \infty$,

$$\begin{aligned} &\chi_\xi(P+l) - \chi_\xi(P) \\ &= NM\log\mathcal{L}_{P+l}(\hat{\theta}_{P+l}) + \xi(P+l)\log NM \\ &- NM\log\mathcal{L}_P(\hat{\theta}_P) - \xi P\log NM \\ &= \xi l\log NM + NM\log\left(1 - \frac{U_l}{NM\mathcal{L}_P(\hat{\theta}_P)} + o\left(\frac{\log NM}{NM}\right)\right) \\ &= \xi l\log NM - \left(\frac{U_l}{\mathcal{L}_P(\hat{\theta}_P)} + o(\log NM)\right)(1 + o(1)) \\ &= \left(\xi l - \frac{U_l}{\mathcal{L}_P(\hat{\theta}_P)\log NM} + o(1)\right)\log NM \ge \\ &\left(\xi l - \frac{l\sup_{\omega,\upsilon} I_w(\omega,\upsilon)}{\mathcal{L}_P(\hat{\theta}_P)\log NM} + o(1)\right)\log NM \\ &= l\left(\xi - \frac{\sup_{\omega,\upsilon} I_w(\omega,\upsilon)}{\sup_{\omega,\upsilon} f_w(\omega,\upsilon)\log NM}\frac{\sup_{\omega,\upsilon} f_w(\omega,\upsilon)}{\mathcal{L}_P(\hat{\theta}_P)} + o(1)\right)\log NM, \end{aligned} \tag{32}$$

where the second equality is obtained by substituting $\mathcal{L}_{P+l}(\hat{\theta}_{P+l})$ using the equality (27). The third equality is due to the property that for $x \to 0$, $\log(1+x) = x(1+o(1))$, where the observation that $\frac{U_l}{NM\mathcal{L}_P(\hat{\theta}_P)} \to 0$ a.s. is due to the boundedness of $\mathcal{L}_P(\hat{\theta}_P)$ from (31) and Assumption 2. The observation that $U_l = O(\log NM)$ follows from [23] (Theorem 1) where it is shown that

$$\limsup_{\Psi(N,M)\to\infty} \frac{\sup_{\omega,\upsilon} I_w(\omega,\upsilon)}{\sup_{\omega,\upsilon} f_w(\omega,\upsilon)\log(NM)} \leq 14 \quad \text{a.s.} \tag{33}$$

Finally, using the triangle inequality it is easy to show that for every pair (ω, υ)

$$f_w(\omega,\upsilon) \leq \sigma^2 \Bigg(\sum_{(r,s)\in D} |a(r,s)| \Bigg)^2. \tag{34}$$

Substituting (31), (33) and (34) into (33) we conclude that

$$\chi_\xi(P+l) - \chi_\xi(P) \geq l\Bigg(\xi - 14\mathcal{A} + o(1)\Bigg)\log NM > 0 \tag{35}$$

for any integer $l \geq 1$ and $\xi > 14\mathcal{A}$. Therefore, a.s. as $\Psi(N,M) \to \infty$, the function $\chi_\xi(k)$ has a **global minimum** for $k = P$. □

The last result generalizes the results of [12] and is similar in its spirit to the result of [19]: On the one hand we preserve the AIC/MDL/BIC form of the model order selection rule. On the other hand, in contrast with the penalty function of AIC and BIC model selection rules, the penalty in (15) is not fixed, but represents a family of penalties, such that they all induce strongly consistent model selection rules. Moreover, it is obvious that the lower bound on ξ depends on the properties of the distribution of the noise field, linearly reflected through the quantity $\mathcal{A}$. It is easy to see that $\mathcal{A} \geq 1$ and equality holds if and only if $a(i,j) = 0$ for all $(i,j) \neq (0,0)$, while $a(0,0) = 1$. In other words, the tightest bound is obtained in the case where the noise field is white.

In general, the problem of finding a tight bound for the parameter ξ remains open. Moreover, we can easily show that by introducing some additional restrictions on the structure of the noise field, we can establish a tighter bound of ξ. We thus modify our earlier Assumption 1, 2 regarding the noise field as follows:

Assumption 1' The noise field $\{w(n,m)\}$ is an infinite order *quarter-plane* MA field, *i.e.*,

$$w(n,m) = \sum_{r,s=0}^{\infty} a(r,s)u(n-r, m-s) \tag{36}$$

where the field $\{u(n,m)\}$ is an i.i.d. real valued zero-mean random field with finite variance σ^2, such that $E[u(n,m)^2 \log|u(n,m)|] < \infty$.

Assumption 2' The sequence $a(i,j)$ is a deterministic sequence which satisfied the condition

$$\sum_{r,s=0}^{\infty} (r+s)|a(r,s)| < \infty. \tag{37}$$

In this case, based on [10], Theorem 3.2 and Assumption 1', 2' we have that

$$\limsup_{\Psi(N,M)\to\infty} \frac{\sup_{\omega,\upsilon} I_w(\omega,\upsilon)}{\sup_{\omega,\upsilon} f_w(\omega,\upsilon)\log(NM)} \leq 8 \quad \text{a.s.} \tag{38}$$

The results of Theorem 1 and 2 are not affected by this assumption. The only change is in Theorem 3. Therefore we can formulate the next theorem:

Theorem 4. *Let Assumptions 1', 2', 3 and 4 be satisfied. Let $\hat{P}$ be given by (15) with $\xi > 8\mathcal{A}$. Then as $\Psi(N,M) \to \infty$*

$$\hat{P} \to P \;\; a.s. \tag{39}$$

The proof of the Theorem 4 is identical to the proof of Theorem 3, where instead of (33) we employ the inequality in (38).

As we have shown, the correct model order is the one for which the global minimum of (15) is obtained and this minimum is the only minimum of (15). Therefore in theory one can terminate the model order selection procedure immediately after discovering the first minimum. Nevertheless, since the LSE is highly non-linear in the sinusoids' parameters and is implemented by non-convex optimization methods which cannot guarantee that the global minimum of the LSE loss function is found, it is advised to proceed with the model order selection procedure for a few more steps after finding a first minimum to ensure that this minimum is indeed the global one. The final result of the model order selection procedure will be the number of sinusoids and their parameters.

5. CONCLUSIONS

We have considered the problem of jointly estimating the number as well as the parameters of two-dimensional sinusoidal signals, observed in the presence of an additive colored noise field. We have established the strong consistency of the LSE when the number of sinusoidal signals is under-estimated, or over-estimated. In the case where the number of sinusoidal signals is under-estimated we have shown the almost sure convergence of the least squares estimates to the parameters of the dominant sinusoids. In the case where this number is over-estimated, the estimated parameter vector obtained by the least squares estimator contains a sub-vector that converges almost surely to the correct parameters of the sinusoids. Based on these results, we proved the strong consistency of a new family of model order selection rules for the number of sinusoidal components and their parameters.

6. REFERENCES

[1] M. Kliger and J. M. Francos, "Strongly Consistent Model Order Selection for Estimating 2-D Sinusoids in Colored Noise," submitted for publication.

[2] H. Akaike, " A New Look at the Statistical Model Identification," *IEEE Trans. Automat. Contr.*, vol. AC-19, pp.716-723, 1974.

[3] K. L. Chung, *A Course in Probability Theory*, Academic Press, 1974.

[4] J. L. Doob, *Stochastic Processes*, John Wiley, 1953.

[5] P. M. Djuric,"A Model Selection Rule for Sinusoids in White Gaussian Noise," *IEEE Trans. Signal Process.*, vol. 44, pp. 1744-1751, 1996.

[6] 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 , pp. 1242-1255, 1996.

[7] J. M. Francos, A. Narashimhan, and J. W. Woods, "Maximum-Likelihood Estimation of Textures Using a Wold Decomposition Model," *IEEE Trans. Image Process.*, vol. 4, pp. 1655-1666, 1995.

[8] J. M. Francos, "Bounds on the Accuracy of Estimating the Parameters of Discrete Homogeneous Random Fields with Mixed Spectral Distributions," *IEEE Trans. Info. Theory*, vol. 43, pp. 908-922, 1997.

[9] E. J. Hannan, " The Estimation of Frequency," *J. Appl. Prob.*,vol. 10 , pp. 510-519, 1973.

[10] S. He, "Uniform Convergency for Weighted Periodogram of Stationary Linear Random Fields," *Chin. Ann. of Math.*, vol. 16B, pp. 331-340, 1995.

[11] S. He, "Parameter Estimation of Hidden Periodic Model in Random Fields," *Science in China (Ser. A)*, vol. 42, pp. 238-245, 1998.

[12] M. Kliger and J. M. Francos, "Strong Consistency of a Family of Model Order Selection Rules for Estimating the Parameters of 2-D Sinusoids in White Noise," *Statistics and Probability Letters*, vol. 78, pp. 3075-3081, 2008.

[13] M. Kliger and J. M. Francos , "Strong Consistency of the Over- and Under-Determined LSE of 2-D Exponentials in White Noise," *IEEE Trans. Info. Theory*, vol. 51, no. 9, pp. 3314-3321, 2005.

[14] D. Kundu and S. Nandi, "Determination of Discrete Spectrum in a Random Field," *Statistica Neerlandica*, vol. 57, pp. 258-283, 2003.

[15] D. Kundu and S. Nandi, " Estimating the Number of Components by a Data Independent Penalty Function Approach," *Journal of Statistical Planning and Inference*, vol. 116, pp. 437-450, 2003.

[16] J. Li and P. Stoica, "Efficient Mixed-Spectrum Estimation with Applications to Target Feature Extraction," *IEEE Trans. Signal Process.*, vol. 44, pp. 281-295, 1996.

[17] J. Li, P.Stoica, and D.Zheng, "An Efficient Algorithm for Tow-dimensional Frequency Estimation," *Multidimensional Systems and Signal Process.*, vol. 7, pp. 151-178, 1996.

[18] A. Prasad, D. Kundu and A. Mitra, " Sequential estimation of two dimensional sinusoidal models," *J. of Probability and Statistical Sciences*, in print.

[19] B. G. Quinn, "Estimating the Number of Terms in a Sinusoidal Regression," *J. Time Series Anal.*, vol. 10, pp. 71-75, 1989.

[20] C. R. Rao, L. Zhao and B. Zhou, "Maximum Likelihood Estimation of 2-D Superimposed Exponential Signals," *IEEE Trans. Signal Process.*, vol. 42 , pp. 1795-1802, 1994.

[21] J. Rissanen, *Stochastic Complexity in Statistical Inquiry*. Singapore: World Scientific, 1989.

[22] G. Schwarz, " Estimating the Dimension of a Model," *Ann. Stat.*, vol. 6, pp. 461-464, 1978.

[23] H. Zhang and V. Mandrekar, "Estimation of Hidden Frequencies for 2D Stationary Processes," *J. Time Ser. Anal.*, vol. 22, pp. 613-629, 2001.

[24] L. C. Zhao, P. R. Krishnaiah and Z. D. Bai, "On Detection of the Number of Signals in Presence of White Noise," *J. Multivariate Anal.*, vol. 20, pp. 1-25, 1986.

[25] L. C. Zhao, P. R. Krishnaiah and Z. D. Bai, "On Detection of the Number of Signals When the Noise Covariance Matrix is Arbitrary," *J. Multivariate Anal.*, vol. 20, pp. 26-49, 1986.

[26] J. Ward, *Space-Time Adaptive Processing for Airborne Radar*, Technical Report 1015, Lincoln Laboratory, Massachusetts Institute of Technology, 1994.

[27] C.F.J. Wu, "Asymptotic Theory of Non-Linear Least Squares Estimation," *Ann. Stat.*, vol. 9, pp. 510-513, 1981.