# Strongly Consistent Model Order Selection for Estimating 2-D Sinusoids in Colored Noise

Mark Kliger and Joseph M. Francos *

December 23, 2012

**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 incorrect. Based on these results, we prove the strong consistency of a new family of model order selection rules.**

**Keywords:** Two-dimensional random fields; model order selection; least squares estimation; strong consistency.

*M. Kliger is with Omek Interactive, Bet Shemesh, Israel, Tel: +972-52-2651810, email: mark.kliger@gmail.com. J. M. Francos is with the Department of Electrical and Computer Engineering, Ben-Gurion University, Beer-Sheva, Israel. Tel: +972-8-6461842, FAX: +972-8-6472949, email: francos@ee.bgu.ac.il

# 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.* [28] 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], [17], [25], 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), [1], the Minimum Description Length (MDL), [23] and the Bayesian Information Criterion (BIC) [24]. 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 functions of both the model order and the log of the size of the data sample. The penalties of MDL and BIC are identical.

In [26] and [27] 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], [16], [18], [19], [21], and the references therein). Quinn, [21], 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 [21].

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.* [20]) or other penalties which were derived for the 1-D case are adopted for the 2-D case (see, *e.g.* [18]).

In [12], following ideas of [21], 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 [12] and [21], 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,

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

where $0 \leq n \leq N-1$, $0 \leq m \leq M-1$ and for each $i$, $\rho_i^0$ 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 the Wold decomposition and is 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

$$(i,j) \succeq (s,t) \text{ iff } (i,j) \in \{(k,l)|k=s, l \geq t\} \cup \{(k,l)|k>s, -\infty \leq l \leq \infty\}. \tag{2}$$

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

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

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.*,

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

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 $\sigma^2$, 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.*,

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

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

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

**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, $\omega_i^0 \neq \omega_j^0$ or $\upsilon_i^0 \neq \upsilon_j^0$, when $i \neq j$.

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 $\Theta_k$ 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 $\rho_i$ is non-zero and bounded for every $1 \le i \le k$ while the pairs $(\omega_i, \upsilon_i)$ are pairwise different, so that no two regressors coincide. We shall refer to $\Theta_k$ 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

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

A vector $\hat{\theta}_k \in \Theta_k$ that minimizes $\mathcal{L}_k(\theta_k)$ is called the *Least Squares Estimate* (LSE). For the case where $k = P$, the asymptotic properties of this estimator are analyzed in detail in [3], under slightly more restrictive assumptions on the properties of the purely-indeterministic component of the observed field, than those made in Assumption 1 and Assumption 2, above. More specifically, it is shown in [3] that if the purely-indeterministic field satisfies a combination of conditions comprised of a strong mixing condition and a condition that it has uniformly bounded $4+\delta$ absolute moments for some $\delta > 0$, the LSE is consistent asymptotically normal (CAN) and asymptotically efficient if the purely-indeterministic field is further assumed to be Gaussian. In [14] it is further shown that under weaker assumptions that those made in the present paper, the LSE is a *strongly consistent* estimator of $\theta_P^0$ and is consistent asymptotically normal.

# 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 Appendix A

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 $\ell_2$ 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 $\sigma^2$, 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 $\Theta_P$, $\Theta_{P+1}$ 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. $\varphi_{per}$ denotes the phase at this point.

**Theorem 2.** *Let Assumptions 1-4 be satisfied. Then, the parameter vector* $\hat{\theta}_{P+1} = (\hat{\rho}_1, \hat{\omega}_1, \hat{\upsilon}_1, \hat{\varphi}_1, \ldots, \hat{\rho}_P, \hat{\omega}_P, \hat{\upsilon}_P, \hat{\varphi}_P, \hat{\rho}_{P+1}, \hat{\omega}_{P+1}, \hat{\upsilon}_{P+1}, \hat{\varphi}_{P+1}) \in \Theta_{P+1}$ *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 Appendix B.

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 $\xi$ 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. Note that in (14) we adopt the general from of the MDL and BIC rules, such that these rules become special cases of (14), where each was derived using different formal reasoning. In this section we provide the conditions for this general form of model order selection rules to be consistent, in the case of the problem of estimating the parameters of 2-D sinusoids in colored noise.

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

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

Let

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

Note that $\mathcal{A}$ is the ratio of the upper bound on the maximal value of the spectral density of the purely-indeterministic field, and its variance. By Assumption 2, $\mathcal{A}$ is bounded.

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 $\Psi(N,M) \to \infty$*

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

*Proof:*

For $k \leq P$,

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

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)

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

From Lemma 3 in Appendix C, we have that as $\Psi(N,M)\to\infty$

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

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 $\Psi(N,M)\to\infty$

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

and similarly

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

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

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

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

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

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

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

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 $\Psi(N, M) \to \infty$,

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

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 [25] (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 $(\omega, \upsilon)$

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

Substituting (31), (33) and (34) into (32) we conclude that

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

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 $k = P$. ■

Note that the condition $\xi > 14\mathcal{A}$ implies that as the peaks of the spectral density of the colored noise become higher, so is the penalty on adding more components to the assumed model.

The last result generalizes the results of [12] and is similar in its spirit to the result of [21]: 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 $\xi$ 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 $\xi$ 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 $\xi$. 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 $\sigma^2$, 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

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

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 $\Psi(N,M) \to \infty$*

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

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 Finite Sample Results

In the following we shall numerically evaluate the performance of the proposed model order selection rule for finite data dimensions. More specifically, we investigate the performance of the proposed model order selection rule as a function of SNR, data dimensions, and for different shapes of the spectral density of the noise field. In these experiments we evaluate the probability of a correct decision by the proposed model order selection rule (14) for different values of $\xi$. The probability of correct decision is defined as the ratio of the number of experiments in which the proposed rule provided the correct model order, normalized by the number of experiments conducted at the specific setting of the experiment. In each setting of the experiment we have conducted 500 Monte-Carlo experiments, *i.e.* , every point in the graphs below was computed by averaging the results over 500 independent trials. The colored noise component of the field

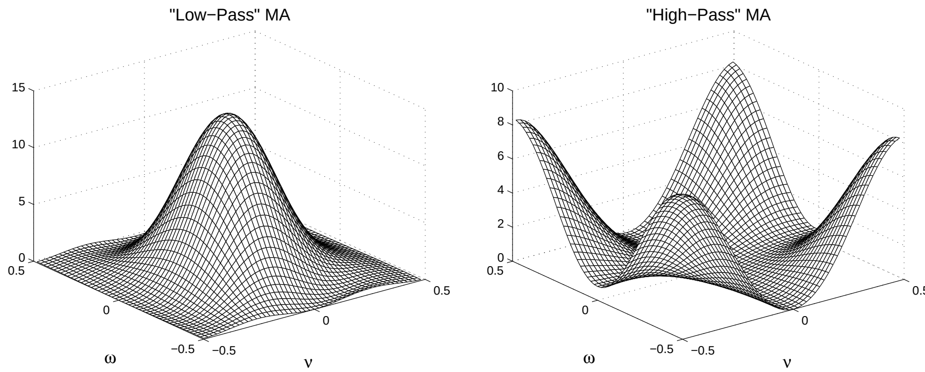


Figure 1: The spectral density function of the "low-pass" (left) and "high-pass" (right) MA models.

is a NSHP MA field with support $D(1,1)$ (see Appendix C for the definition of a finite support $D(k,l)$). To investigate the dependence of the performance of the proposed model order selection rule on the shape of the spectral density of the colored noise, the test is performed for both "low-pass" type MA model and for a "high-pass" type MA model.

In the first example the noise field is a NSHP MA "high-pass" field. The MA model parameters are $a(0,1) = -0.9$, $a(1,-1) = 0.1$, $a(1,0) = -0.5$, $a(1,1) = 0.4$. In the second part of this example the colored noise component of the field is a NSHP MA "low-pass" field. The MA model parameters are $a(0,1) = 0.9$, $a(1,-1) = 0.4$, $a(1,0) = 0.8$, $a(1,1) = 0.6$. Thus, $\mathcal{A} = 2.395$ for the "high-pass" MA noise field and $\mathcal{A} = 3.7005$ for the "low-pass" field. The driving noise of the MA model, in both cases, is a zero mean, white noise field with a unit variance. The spectral density functions of these fields are depicted in Fig. 1. The harmonic component of the field has three ($P = 3$) sinusoidal components with frequencies $(0.1, 0.2)$, $(-0.3, 0.4)$ and $(0.1 + 1/N, 0.2 + 1/M)$ of equal amplitudes. The SNR in the case of these examples in defined as the ratio of the sinusoidal power to that of the driving noise of the MA model. We test the model order selection rule for different data sizes such that $N, M = 32, 64, 128, 256, 512$.

Figure 2 provides the probability of correct order selection, (*i.e.* , $\hat{P} = 3$) as a function of the data dimensions, for different values of $\xi$, and a fixed SNR of -15dB, for the two different models of the colored noise component. It is concluded that for all values of $\xi$ around the minimal value determined by the consistency requirement, and for data dimensions that are higher than $128 \times 128$ the probability of correct decision is 1. Moreover, as shown in Fig. 3, and Fig. 4 already for relatively small dimensions of the observed field ($64 \times 64$) and low SNR values, the probability of correct decision is 1, for $\xi$ values around the minimal value of $14\mathcal{A}$ predicted by the consistency constraint: For $\xi = 8\mathcal{A}$ to $\xi = 20\mathcal{A}$ we observe the same "threshold effect" as a function of both SNR and the rate in which the penalty term increases as expressed by $\xi$: While increasing $\xi$ for

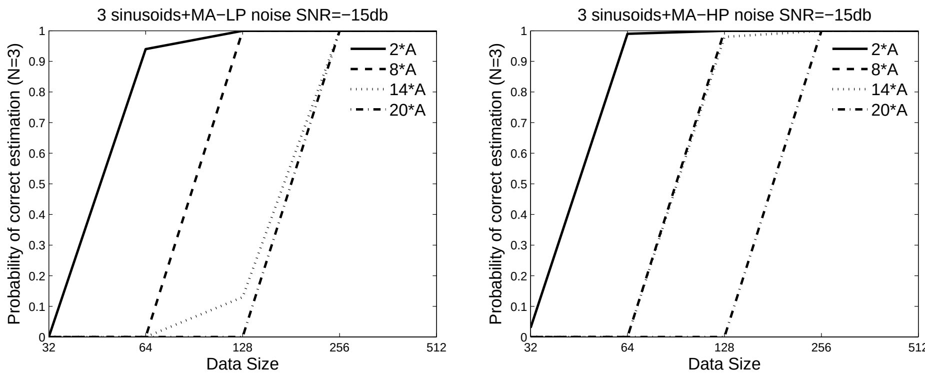


Figure 2: Probability of correct order selection, ($i.e.$ , $\hat{P} = 3$) as a function of the data dimensions, for different values of $\xi$, and a fixed SNR of -15dB.

a given SNR results in a larger number of underestimations, when we increase the SNR, or data dimensions, for a fixed value of $\xi$ there is a threshold point where the probability becomes 1 again. In the case of low SNR incorrect model-order selection is due to overestimation of the model order. Obviously, for extremely high values of $\xi$, the proposed rule results in underestimation of the model order even for relativity high SNR's, due to the extremely high weight given to the penalty term. On the other hand, for low $\xi$ values, such as where $\xi = 2\mathcal{A}$ the errors in selecting the model order are due to overestimation. Note however, that as data dimensions increase (we approach the asymptotic assumptions on data dimensions) the model order selection rule is becoming consistent for lower SNR values and for all $\xi$ values around the minimal value of $14\mathcal{A}$ predicted by the consistency constraint. Also note that for all data dimensions being considered in these experiment, for $\xi = 2\mathcal{A}$ (which is far from the minimal value of $\xi = 14\mathcal{A}$ required to guarantee consistency) the probability of correct order selection drops sharply as the SNR is getting higher due to overestimation of the model order. This is clearly a result of the lower significance of the penalty term.

Furthermore, the experimental results indicate that the performance of the proposed model order selection rule is similar for both types of spectral density models being considered.

# 6 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

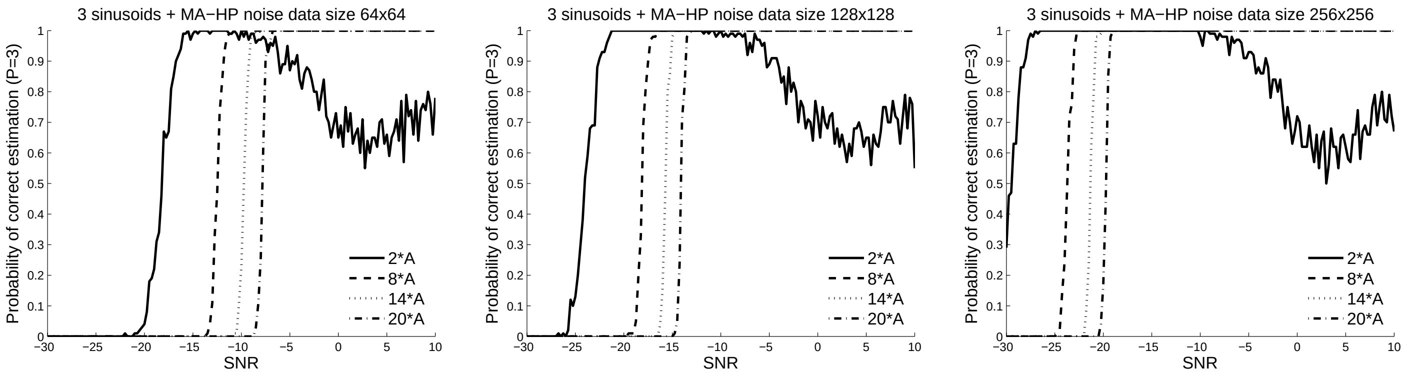


Figure 3: Probability of correct order selection, (*i.e.* , $\hat{P} = 3$) as a function of the SNR, for different values of $\xi$. The observed field dimensions are (from left to right) $64 \times 64$, $128 \times 128$, and $256 \times 256$ samples. The colored noise component of the field is a NSHP MA "high-pass" field.

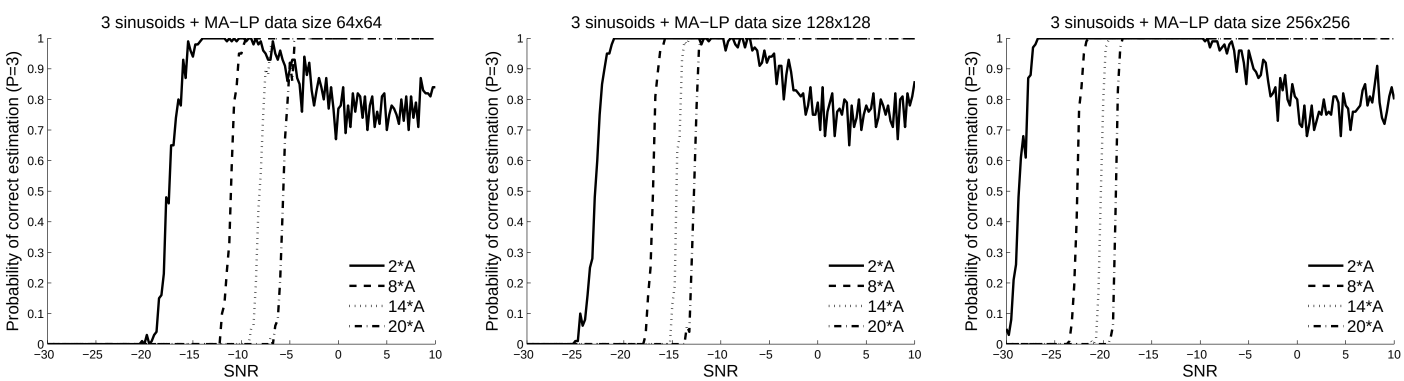


Figure 4: Probability of correct order selection, (*i.e.* , $\hat{P} = 3$) as a function of the SNR, for different values of $\xi$. The observed field dimensions are (from left to right) $64 \times 64$, $128 \times 128$, and $256 \times 256$ samples. The colored noise component of the field is a NSHP MA "low-pass" field.

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. The applicability and validity of the asymptotic results to finite dimensional observations is demonstrated using Monte-Carlo experiments.

## Appendix A: Proof of Theorem 1

The proof of the Theorem 1 follows similar lines to those of the proof of Theorem 1 [13] where we considered a less general case in which the observed field is composed of complex 2-D exponentials in an additive white noise field.

In order to prove Theorem 1 we have to establish some auxiliary results.

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

$$\Delta_\delta = \left[\bigcup_{i=1}^{k} \mathcal{R}_{i\delta}\right] \cup \left[\bigcup_{i=1}^{k} \Phi_{i\delta}\right] \cup \left[\bigcup_{i=1}^{k} W_{i\delta}\right] \cup \left[\bigcup_{i=1}^{k} V_{i\delta}\right] \ , \tag{40}$$

where

$$\begin{aligned} \mathcal{R}_{i\delta} &= \left\{\theta_k \in \Theta_k : \ |\rho_i - \rho_i^0| \geq \delta; \delta > 0 \ \right\} \ , \\ \Phi_{i\delta} &= \left\{\theta_k \in \Theta_k : \ |\varphi_i - \varphi_i^0| \geq \delta; \delta > 0 \ \right\} \ , \\ W_{i\delta} &= \left\{\theta_k \in \Theta_k : \ |\omega_i - \omega_i^0| \geq \delta; \delta > 0 \ \right\} \ , \\ V_{i\delta} &= \left\{\theta_k \in \Theta_k : \ |\upsilon_i - \upsilon_i^0| \geq \delta; \delta > 0 \ \right\} \ . \end{aligned} \tag{41}$$

The next lemma shows that the true parameters of the $k$ dominant sinusoids of the model (1) asymptotically minimize the $k$th-order least squares function (7).

**Lemma 1.**

$$\liminf_{\Psi(N,M)\to\infty} \inf_{\theta_k \in \Delta_\delta} \left(\mathcal{L}_k(\theta_k) - \mathcal{L}_k(\theta_k^0)\right) > 0 \quad a.s. \tag{42}$$

*Proof:* In the following we first show that on $\Delta_\delta$ the sequence $\mathcal{L}_k(\theta_k) - \mathcal{L}_k(\theta_k^0)$ (indexed in $N, M$) is uniformly lower bounded by a strictly positive constant as $\Psi(N,M) \to \infty$. Since the sequence elements are uniformly lower bounded by a strictly positive constant, the sequence of infimums, $\inf\limits_{\theta_k \in \Delta_\delta} (\mathcal{L}_k(\theta_k) - \mathcal{L}_k(\theta_k^0))$, is uniformly lower bounded by the same strictly positive constant as $\Psi(N, M) \to \infty$. Hence, $\liminf\limits_{\Psi(N,M)\to\infty} \inf\limits_{\theta_k \in \Delta_\delta} (\mathcal{L}_k(\theta_k) - \mathcal{L}_k(\theta_k^0))$ is also lower bounded by the same constat.

Thus, we first prove that the sequence $\mathcal{L}_k(\theta_k) - \mathcal{L}_k(\theta_k^0)$ is uniformly lower bounded away from zero on $\Delta_\delta$ as $\Psi(N, M) \to \infty$.

$$
\begin{aligned}
&\mathcal{L}_k(\theta_k) - \mathcal{L}_k(\theta_k^0) \\
&= \tfrac{1}{NM} \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} \left( y(n,m) - \sum_{i=1}^{k} \rho_i \cos(\omega_i n + \upsilon_i m + \varphi_i) \right)^2 \\
&- \tfrac{1}{NM} \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} \left( y(n,m) - \sum_{i=1}^{k} \rho_i^0 \cos(\omega_i^0 n + \upsilon_i^0 m + \varphi_i^0) \right)^2 \\
&= \tfrac{1}{NM} \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} \left( \sum_{i=1}^{P} \rho_i^0 \cos(\omega_i^0 n + \upsilon_i^0 m + \varphi_i^0) + w(n,m) - \sum_{i=1}^{k} \rho_i \cos(\omega_i n + \upsilon_i m + \varphi_i) \right)^2 \\
&- \tfrac{1}{NM} \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} \left( \sum_{i=k+1}^{P} \rho_i^0 \cos(\omega_i^0 n + \upsilon_i^0 m + \varphi_i^0) + w(n,m) \right)^2 \\
&= \tfrac{1}{NM} \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} \left( \sum_{i=1}^{k} \rho_i^0 \cos(\omega_i^0 n + \upsilon_i^0 m + \varphi_i^0) - \sum_{i=1}^{k} \rho_i \cos(\omega_i n + \upsilon_i m + \varphi_i) \right)^2 \\
&+ \tfrac{2}{NM} \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} \left( \sum_{i=k+1}^{P} \rho_i^0 \cos(\omega_i^0 n + \upsilon_i^0 m + \varphi_i^0) \right) \\
&\left( \sum_{i=1}^{k} \rho_i^0 \cos(\omega_i^0 n + \upsilon_i^0 m + \varphi_i^0) - \sum_{i=1}^{k} \rho_i \cos(\omega_i n + \upsilon_i m + \varphi_i) \right) \\
&+ \tfrac{2}{NM} \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} w(n,m) \left( \sum_{i=1}^{k} \rho_i^0 \cos(\omega_i^0 n + \upsilon_i^0 m + \varphi_i^0) - \sum_{i=1}^{k} \rho_i \cos(\omega_i n + \upsilon_i m + \varphi_i) \right) \\
&= I_1 + I_2 + I_3.
\end{aligned}
\tag{43}
$$

Thus, to check the asymptotic behavior of L.H.S. of (43) we have to evaluate $\lim\limits_{\Psi(N,M)\to\infty} (I_1 + I_2 + I_3)$ for all vectors $\theta_k \in \Delta_\delta$:

$$
\begin{aligned}
&\lim_{\Psi(N,M)\to\infty} I_1 = \lim_{\Psi(N,M)\to\infty} \tfrac{1}{NM} \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} \left( \sum_{i=1}^{k} \rho_i^0 \cos(\omega_i^0 n + \upsilon_i^0 m + \varphi_i^0) \right)^2 \\
&- \lim_{\Psi(N,M)\to\infty} \left[ 2 \tfrac{1}{NM} \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} \sum_{i=1}^{k} \sum_{j=1}^{k} \rho_i \rho_j^0 \cos(\omega_i n + \upsilon_i m + \varphi_i) \cos(\omega_j^0 n + \upsilon_j^0 m + \varphi_j^0) \right] \\
&+ \lim_{\Psi(N,M)\to\infty} \tfrac{1}{NM} \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} \left( \sum_{i=1}^{k} \rho_i \cos(\omega_i n + \upsilon_i m + \varphi_i) \right)^2 = T_1 + T_2 + T_3.
\end{aligned}
\tag{44}
$$

Recall that for $|\rho| < \infty$ and $\varphi \in [0, 2\pi)$

$$
\lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \rho \cos(\omega n + \varphi) = 0, \tag{45}
$$

uniformly in $\omega$ on any closed interval in $(0,2\pi)$. The same equality holds for the sine function. Hence,

$$\begin{aligned}T_3 &= \lim_{\Psi(N,M)\to\infty} \tfrac{1}{NM}\sum_{n=0}^{N-1}\sum_{m=0}^{M-1}\left(\sum_{i=1}^{k}\rho_i\cos(\omega_i n+\upsilon_i m+\varphi_i)\right)^2 = \sum_{i=1}^{k}\tfrac{(\rho_i)^2}{2}\\ &+ \lim_{\Psi(N,M)\to\infty}\tfrac{1}{NM}\sum_{n=0}^{N-1}\sum_{m=0}^{M-1}\sum_{\substack{i=1\\ i\neq j}}^{k}\sum_{j=1}^{k}\rho_i\rho_j\cos(\omega_i n+\upsilon_i m+\varphi_i)\cos(\omega_j n+\upsilon_j m+\varphi_j).\end{aligned}\tag{46}$$

Since the pairs $(\omega_i,\upsilon_i)$ are pairwise different, then on any closed interval in $(0,2\pi)$ the sequence of partial sums $\frac{1}{NM}\sum\limits_{n=0}^{N-1}\sum\limits_{m=0}^{M-1}\sum\limits_{\substack{i=1\\ i\neq j}}^{k}\sum\limits_{j=1}^{k}\rho_i\rho_j\cos(\omega_i n+\upsilon_i m+\varphi_i)\cos(\omega_j n+\upsilon_j m+\varphi_j)$ converges uniformly to zero as $\Psi(N,M)\to\infty$.

Hence,

$$T_3=\sum_{i=1}^{k}\tfrac{(\rho_i)^2}{2},\tag{47}$$

as $\Psi(N,M)\to\infty$ uniformly on $\Delta_\delta$. Similarly,

$$T_1=\lim_{\Psi(N,M)\to\infty}\frac{1}{NM}\sum_{n=0}^{N-1}\sum_{m=0}^{M-1}\left(\sum_{i=1}^{k}\rho_i^0\cos(\omega_i^0 n+\upsilon_i^0 m+\varphi_i^0)\right)^2=\sum_{i=1}^{k}\frac{(\rho_i^0)^2}{2},\tag{48}$$

independently of $\theta_k$.

Leaving $T_2$ unchanged we obtain

$$\begin{aligned}&\lim_{\Psi(N,M)\to\infty} I_1=\sum_{i=1}^{k}\left(\tfrac{(\rho_i^0)^2}{2}+\tfrac{(\rho_i)^2}{2}\right)\\ &-\lim_{\Psi(N,M)\to\infty}\tfrac{2}{NM}\sum_{n=0}^{N-1}\sum_{m=0}^{M-1}\sum_{i=1}^{k}\sum_{j=1}^{k}\rho_i\rho_j^0\cos(\omega_i n+\upsilon_i m+\varphi_i)\cos(\omega_j^0 n+\upsilon_j^0 m+\varphi_j^0),\end{aligned}\tag{49}$$

uniformly on $\Delta_\delta$.

Using the similar considerations to those employed in the evaluation of (48) we obtain

$$\begin{aligned}&\lim_{\Psi(N,M)\to\infty} I_2=\lim_{\Psi(N,M)\to\infty}\Bigg[\tfrac{2}{NM}\sum_{n=0}^{N-1}\sum_{m=0}^{M-1}\left(\sum_{i=k+1}^{P}\rho_i^0\cos(\omega_i^0 n+\upsilon_i^0 m+\varphi_i^0)\right)\\ &\left(\sum_{i=1}^{k}\rho_i^0\cos(\omega_i^0 n+\upsilon_i^0 m+\varphi_i^0)-\sum_{i=1}^{k}\rho_i\cos(\omega_i n+\upsilon_i m+\varphi_i)\right)\Bigg]\\ &=-\lim_{\Psi(N,M)\to\infty}\left[\tfrac{2}{NM}\sum_{n=0}^{N-1}\sum_{m=0}^{M-1}\sum_{i=1}^{k}\sum_{j=k+1}^{P}\rho_i\rho_j^0\cos(\omega_i n+\upsilon_i m+\varphi_i)\cos(\omega_j^0 n+\upsilon_j^0 m+\varphi_j^0)\right].\end{aligned}\tag{50}$$

By Lemma 3 in Appendix C, we have that a.s. as $\Psi(N,M)\to\infty$ :

$$\sup_{\theta_k\in\Delta_\delta}\left|\frac{2}{NM}\sum_{n=0}^{N-1}\sum_{m=0}^{M-1}w(n,m)\Bigg(\sum_{i=1}^{k}\rho_i^0\cos(\omega_i^0 n+\upsilon_i^0 m+\varphi_i^0)-\sum_{i=1}^{k}\rho_i\cos(\omega_i n+\upsilon_i m+\varphi_i)\Bigg)\right|\to 0.\tag{51}$$

Hence $I_3 \to 0$ $a.s.$ as $\Psi(N,M) \to \infty$ uniformly on $\Delta_\delta$. Using (49), (50) and (51) we conclude that a.s.

$$\begin{aligned}
&\lim_{\Psi(N,M)\to\infty} (\mathcal{L}_k(\theta_k) - \mathcal{L}_k(\theta_k^0)) = \sum_{i=1}^{k} \left( \tfrac{(\rho_i^0)^2}{2} + \tfrac{(\rho_i)^2}{2} \right) \\
&- \lim_{\Psi(N,M)\to\infty} \tfrac{2}{NM} \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} \sum_{i=1}^{k} \sum_{j=1}^{P} \rho_i \rho_j^0 \cos(\omega_i n + \upsilon_i m + \varphi_i) \cos(\omega_j^0 n + \upsilon_j^0 m + \varphi_j^0).
\end{aligned} \tag{52}$$

To complete the evaluation of (52) we consider the vectors $\theta_k \in \Delta_\delta$. Let us first assume that $\Delta_\delta \equiv \mathcal{R}_{q\delta}$ for some $q$, $1 \le q \le k$. Thus, the coordinate $\rho_q$ of each vector in this subset is different from the corresponding coordinate $\rho_q^0$ by at least $\delta > 0$. Consider first the case where all the other elements of the vector $\theta_k \in \mathcal{R}_{q\delta}$ are identical to the corresponding elements of $\theta_k^0$. Since by this assumption $\omega_j = \omega_j^0$, $\upsilon_j = \upsilon_j^0$, $\varphi_j = \varphi_j^0$ for $1 \le j \le k$, and $\rho_j = \rho_j^0$ for $1 \le j \le k$, $j \ne q$, on this set we have

$$\begin{aligned}
&\lim_{\Psi(N,M)\to\infty} \big(\mathcal{L}_k(\theta_k) - \mathcal{L}_k(\theta_k^0)\big) = \left( \frac{\rho_q^0}{\sqrt{2}} - \frac{\rho_q}{\sqrt{2}} \right)^2 \\
&- \lim_{\Psi(N,M)\to\infty} \frac{2}{NM} \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} \sum_{\substack{i=1 \\ i\ne j}}^{k} \sum_{j=1}^{P} \rho_i \rho_j^0 \cos(\omega_i n + \upsilon_i m + \varphi_i) \cos(\omega_j^0 n + \upsilon_j^0 m + \varphi_j^0) \\
&= \left( \frac{\rho_q^0}{\sqrt{2}} - \frac{\rho_q}{\sqrt{2}} \right)^2 \ge \frac{\delta^2}{2} > 0,
\end{aligned} \tag{53}$$

uniformly in $\rho_q$, where the second equality is due to Assumption 3 and following the arguments employed to obtain (47).

Assume next that $\theta_k \in \mathcal{R}_{q\delta}$ ($i.e.$, the coordinate $\rho_q$ is different from the corresponding coordinate $\rho_q^0$ by at least $\delta > 0$) and that in addition, there exists an element $\rho_t$ of $\theta_k$, such that $1 \le t \le k$, $t \ne q$ and $|\rho_t - \rho_t^0| \ge \lambda, \lambda > 0$ while all the other elements of the vector $\theta_k$ are identical to the corresponding elements of $\theta_k^0$. Following a similar derivation to the one in (53) we conclude that

$$\lim_{\Psi(N,M)\to\infty} \big(\mathcal{L}_k(\theta_k) - \mathcal{L}_k(\theta_k^0)\big) = \left( \frac{\rho_q^0}{\sqrt{2}} - \frac{\rho_q}{\sqrt{2}} \right)^2 + \left( \frac{\rho_t^0}{\sqrt{2}} - \frac{\rho_t}{\sqrt{2}} \right)^2 \ge \frac{\delta^2}{2} + \frac{\lambda^2}{2} > \frac{\delta^2}{2}, \tag{54}$$

uniformly in $\rho_q$ and $\rho_t$.

Consider the case where $\theta_k \in \mathcal{R}_{q\delta}$ while there exists an element $\varphi_l$ of $\theta_k \in \mathcal{R}_{q\delta}$, such that $|\varphi_l - \varphi_l^0| \ge \eta, \eta > 0$ and all the other elements of the vector $\theta_k$ are identical to the corresponding elements of $\theta_k^0$. Following a similar derivation to the one in (53) we conclude that

$$\lim_{\Psi(N,M)\to\infty} \big(\mathcal{L}_k(\theta_k) - \mathcal{L}_k(\theta_k^0)\big) = \left\{ \begin{array}{ll} \left( \frac{\rho_q^0}{\sqrt{2}} - \frac{\rho_q}{\sqrt{2}} \right)^2 + (\rho_l^0)^2 - (\rho_l^0)^2 \cos(\varphi_l - \varphi_l^0), & l \ne q \\ \frac{(\rho_q^0)^2}{2} + \frac{(\rho_q)^2}{2} - \rho_q^0 \rho_q \cos(\varphi_q - \varphi_q^0), & l = q \end{array} \right. > \frac{\delta^2}{2}, \tag{55}$$

uniformly in $\rho_q$ and $\varphi_l$.

Finally, consider the case where $\theta_k \in \mathcal{R}_{q\delta}$ while there exists an element $\omega_l$ of $\theta_k \in \mathcal{R}_{q\delta}$, such that $|\omega_l - \omega_l^0| \geq \eta, \eta > 0$ and all the other elements of the vector $\theta_k$ are identical to the corresponding elements of $\theta_k^0$. Following a similar derivation to the one in (53) we conclude that

$$\liminf_{\Psi(N,M)\to\infty} \left(\mathcal{L}_k(\theta_k) - \mathcal{L}_k(\theta_k^0)\right) = \begin{cases} \left(\frac{\rho_q^0}{\sqrt{2}} - \frac{\rho_q}{\sqrt{2}}\right)^2 + (\rho_l^0)^2, & l \neq q \\ \frac{(\rho_q^0)^2}{2} + \frac{(\rho_q)^2}{2}, & l = q \end{cases} > \frac{\delta^2}{2}, \tag{56}$$

uniformly in $\rho_q$ and $\omega_l$.

From the above analysis it is clear that $\lim\limits_{\Psi(N,M)\to\infty} (\mathcal{L}_k(\theta_k) - \mathcal{L}_k(\theta_k^0))$ is lower bounded by $\frac{\delta^2}{2}$ uniformly in $\mathcal{R}_{q\delta}$.

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

$$\Lambda = \left\{\theta_k \in W_{q\delta} \cup V_{q\delta} : \ \exists p,\ k+1 \leq p \leq P,\ (\omega_q, \upsilon_q) = (\omega_p^0, \upsilon_p^0) \ \right\} \subset W_{q\delta} \cup V_{q\delta} \tag{57}$$

This subset includes vectors in $\Theta_k$, such that their coordinate pairs $(\omega_q, \upsilon_q)$ are different from the corresponding pairs of $\theta_k^0$ and equal to some pair $(\omega_p^0, \upsilon_p^0)$ where $p \geq k+1$. As above, the minimum is obtained when all the other elements of $\theta_k$ are identical to the corresponding elements of $\theta_k^0$. Hence, uniformly on $\Lambda$, we have

$$\begin{aligned} &\lim_{\Psi(N,M)\to\infty} (\mathcal{L}_k(\theta_k) - \mathcal{L}_k(\theta_k^0)) \geq \tfrac{(\rho_q^0)^2}{2} + \tfrac{(\rho_q)^2}{2} - \rho_p^0\rho_q \\ &= \tfrac{(\rho_q^0)^2}{2} - \tfrac{(\rho_p^0)^2}{2} + \left(\tfrac{\rho_p^0}{\sqrt{2}} - \tfrac{\rho_q}{\sqrt{2}}\right)^2 \geq \tfrac{(\rho_q^0)^2}{2} - \tfrac{(\rho_p^0)^2}{2} = \epsilon_\Lambda > 0, \end{aligned} \tag{58}$$

where the last inequality is due to Assumption 4.

On the complementary set:

$$\Lambda^c = (W_{q\delta} \cup V_{q\delta}) \setminus \Lambda = \left\{\theta_k \in W_{q\delta} \cup V_{q\delta} : \ (\omega_q, \upsilon_q) \neq (\omega_p^0, \upsilon_p^0),\ \forall p,\ k+1 \leq p \leq P \ \right\} \tag{59}$$

we have

$$\lim_{\Psi(N,M)\to\infty} (\mathcal{L}_k(\theta_k) - \mathcal{L}_k(\theta_k^0)) \geq \tfrac{(\rho_q^0)^2}{2} + \tfrac{(\rho_q)^2}{2} \geq \tfrac{(\rho_q^0)^2}{2} = \epsilon_{\Lambda^c} > 0. \tag{60}$$

Finally, on the set $\Phi_{q\delta}$, the coordinate $\varphi_q$ of each vector in this subset is different from the corresponding coordinate $\varphi_q^0$ by at least $\delta > 0$. As in previous cases , the minimum is obtained when all the other elements of $\theta_k \in \Phi_{q\delta}$ are identical to the corresponding elements of $\theta_k^0$. Hence, uniformly on $\Phi_{q\delta}$, we have

$$\lim_{\Psi(N,M)\to\infty} (\mathcal{L}_k(\theta_k) - \mathcal{L}_k(\theta_k^0)) \geq (\rho_q^0)^2 - (\rho_q^0)^2\cos(\varphi_q - \varphi_q^0) \geq (\rho_q^0)^2(1-\cos\delta) = \epsilon_{\Phi_{q\delta}} > 0. \tag{61}$$

Let $\epsilon_q = \min(\frac{\delta^2}{2}, \epsilon_\Lambda, \epsilon_{\Lambda^c}, \epsilon_{\Phi_{q\delta}})$. Collecting (53),(58), (60) and (61) together we conclude that the sequence $\mathcal{L}_k(\theta_k) - \mathcal{L}_k(\theta_k^0)$ is lower bounded by $\epsilon_q > 0$ uniformly on $\mathcal{R}_{q\delta} \cup \Phi_{q\delta} \cup W_{q\delta} \cup V_{q\delta}$ as $\Psi(N,M) \to \infty$.

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

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

$$\inf_{\theta_k\in\Delta_\delta} \left(\mathcal{L}_k(\theta_k) - \mathcal{L}_k(\theta_k^0)\right) \geq \epsilon, \tag{62}$$

as $\Psi(N,M) \to \infty$.

Hence, by the definition of lim inf

$$\liminf_{\Psi(N,M)\to\infty} \inf_{\theta_k\in\Delta_\delta} \left(\mathcal{L}_k(\theta_k) - \mathcal{L}_k(\theta_k^0)\right) \geq \epsilon > 0. \tag{63}$$

■

**Remark**: Lemma 1 (and consequently Theorem 1) remain 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 $\sigma^2$, and the sequence $a(i,j)$ is a square summable deterministic sequence, *i.e.*, $\sum_{(r,s)\in D} a^2(r,s) < \infty$, then Lemma 1 and Theorem 1 hold.

To proceed to the proof of Theorem 1 we need the next result:

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

$$\Pr\{x_n \leq 0 \ i.o.\} \leq \Pr\{\liminf_{n\to\infty} x_n \leq 0\} \tag{64}$$

**Proof:** See [13].

Combining the above lemmas we will be able to prove Theorem 1.

*Proof:* (Theorem 1) The proof follows an argument proposed by Wu [29], Lemma 1. Let $\hat{\theta}_k = (\hat{\rho}_1, \hat{\omega}_1, \hat{\upsilon}_1, \hat{\varphi}_1, \ldots, \hat{\rho}_k, \hat{\omega}_k, \hat{\upsilon}_k, \hat{\varphi}_k)$ be a parameter vector that minimizes (7). Assume that the proposition $\hat{\theta}_k \to \theta_k^0 \;\; a.s. \;\; as \;\; \Psi(N,M) \to \infty$ is not true. Then, there exists some $\delta > 0$, such that ([2], Theorem 4.2.2, p. 69),

$$\Pr(\hat{\theta}_k \in \Delta_\delta \;\; i.o.) > 0. \tag{65}$$

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

$$\Pr(\inf_{\theta_k \in \Delta_\delta} \big(\mathcal{L}_k(\theta_k) - \mathcal{L}_k(\theta_k^0)\big) \le 0 \;\; i.o.) > 0. \tag{66}$$

Using Lemma 2 we obtain

$$\Pr(\liminf_{\Psi(N,M)\to\infty} \inf_{\theta_k \in \Delta_\delta} \big(\mathcal{L}_k(\theta_k) - \mathcal{L}_k(\theta_k^0)\big) \le 0) \ge \Pr(\inf_{\theta_k \in \Delta_\delta} \big(\mathcal{L}_k(\theta_k) - \mathcal{L}_k(\theta_k^0)\big) \le 0 \;\; i.o.) > 0, \tag{67}$$

which contradicts (42) from Lemma 1. Hence,

$$\hat{\theta}_k \to \theta_k^0 \;\; a.s. \;\; as \;\; \Psi(N,M) \to \infty. \tag{68}$$

$\blacksquare$

## Appendix B: Proof of Theorem 2

The proof of the Theorem 2 follows similar lines to those of the proof of Theorem 2 [13] where we considered a less general case in which the observed field is composed of complex 2-D exponentials in an additive white noise field.

*Proof:* Let $\theta_{P+1} = (\rho_1, \omega_1, \upsilon_1, \varphi_1, \ldots, \rho_P, \omega_P, \upsilon_P, \varphi_P, \rho_{P+1}, \omega_{P+1}, \upsilon_{P+1}, \varphi_{P+1})$, be some vector in the parameter space $\Theta_{P+1}$. We have,

$$\begin{aligned}
&\mathcal{L}_{P+1}(\theta_{P+1}) = \tfrac{1}{NM} \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} \left( y(n,m) - \sum_{i=1}^{P+1} \rho_i \cos(\omega_i n + \upsilon_i m + \varphi_i) \right)^2 \\
&= \tfrac{1}{NM} \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} \left( y(n,m) - \sum_{i=1}^{P} \rho_i \cos(\omega_i n + \upsilon_i m + \varphi_i) \right)^2 \\
&+ \tfrac{1}{NM} \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} \left( \rho_{P+1} \cos(\omega_{P+1} n + \upsilon_{P+1} m + \varphi_{P+1}) \right)^2 \\
&- \tfrac{2}{NM} \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} \left( y(n,m) - \sum_{i=1}^{P} \rho_i \cos(\omega_i n + \upsilon_i m + \varphi_i) \right) \left( \rho_{P+1} \cos(\omega_{P+1} n + \upsilon_{P+1} m + \varphi_{P+1}) \right) \\
&= \mathcal{L}_P(\theta_P) + \tfrac{\rho_{P+1}^2}{2} + \tfrac{1}{2NM} \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} \rho_{P+1}^2 \cos(2\omega_{P+1} n + 2\upsilon_{P+1} m + 2\varphi_{P+1}) \\
&- \tfrac{2}{NM} \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} w(n,m) \rho_{P+1} \cos(\omega_{P+1} n + \upsilon_{P+1} m + \varphi_{P+1}) \\
&- \tfrac{2}{NM} \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} \left( \sum_{i=1}^{P} \rho_i^0 \cos(\omega_i^0 n + \upsilon_i^0 m + \varphi_i^0) - \sum_{i=1}^{P} \rho_i \cos(\omega_i n + \upsilon_i m + \varphi_i) \right) \\
&\left( \rho_{P+1} \cos(\omega_{P+1} n + \upsilon_{P+1} m + \varphi_{P+1}) \right) = H_1(\theta_{P+1}) + H_2(\theta_{P+1}) + H_3(\theta_{P+1})
\end{aligned} \tag{69}$$

where, $\theta_P = (\rho_1, \omega_1, \upsilon_1, \varphi_1, \ldots, \rho_P, \omega_P, \upsilon_P, \varphi_P) \in \Theta_P$ and,

$$H_1(\theta_{P+1}) = \mathcal{L}_P(\rho_1, \omega_1, \upsilon_1, \varphi_1, \ldots, \rho_P, \omega_P, \upsilon_P, \varphi_P) = \mathcal{L}_P(\theta_P), \tag{70}$$

$$H_2(\theta_{P+1}) = \frac{\rho_{P+1}^2}{2} - \frac{2}{NM}\sum_{n=0}^{N-1}\sum_{m=0}^{M-1} w(n,m)\rho_{P+1}\cos(\omega_{P+1}n + \upsilon_{P+1}m + \varphi_{P+1}), \tag{71}$$

$$\begin{aligned}
H_3(\theta_{P+1}) &= \frac{1}{2NM}\sum_{n=0}^{N-1}\sum_{m=0}^{M-1} \rho_{P+1}^2 \cos(2\omega_{P+1}n + 2\upsilon_{P+1}m + 2\varphi_{P+1}) \\
&- \frac{2}{NM}\sum_{n=0}^{N-1}\sum_{m=0}^{M-1}\Bigg(\sum_{i=1}^{P}\rho_i^0\cos(\omega_i^0 n + \upsilon_i^0 m + \varphi_i^0) - \sum_{i=1}^{P}\rho_i\cos(\omega_i n + \upsilon_i m + \varphi_i)\Bigg) \\
&\Bigg(\rho_{P+1}\cos(\omega_{P+1}n + \upsilon_{P+1}m + \varphi_{P+1})\Bigg).
\end{aligned} \tag{72}$$

Let $\hat{\theta}_P = (\hat{\rho}_1, \hat{\omega}_1, \hat{\upsilon}_1, \hat{\varphi}_1, \ldots, \hat{\rho}_P, \hat{\omega}_P, \hat{\upsilon}_P, \hat{\varphi}_P)$ be a vector in $\Theta_P$ that minimizes $H_1(\theta_{P+1}) = \mathcal{L}_P(\theta_P)$. From [14] (or using Theorem 1 in the previous section),

$$\hat{\theta}_P \to \theta_P^0 \quad a.s. \quad \text{as} \quad \Psi(N,M) \to \infty. \tag{73}$$

The function $H_2$ is a function of $\rho_{P+1}, \omega_{P+1}, \upsilon_{P+1}, \varphi_{P+1}$ 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 $\rho^e, \omega^e, \upsilon^e, \varphi^e$ denote an extremum point of $H_2$. Then at this point

$$H_2(\rho^e, \omega^e, \upsilon^e, \varphi^e) = -\frac{I_w(\omega^e, \upsilon^e)}{NM}. \tag{74}$$

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

$$(\hat{\omega}_{P+1}, \hat{\upsilon}_{P+1}) = \underset{(\omega,\upsilon)\in(0,2\pi)^2}{\arg\min}\ H_2(\rho_{P+1}, \omega_{P+1}, \upsilon_{P+1}, \varphi_{P+1}) = \underset{(\omega,\upsilon)\in(0,2\pi)^2}{\arg\max}\ I_w(\omega, \upsilon), \tag{75}$$

and

$$\hat{\rho}_{P+1}^2 = \frac{2}{NM} I_w(\hat{\omega}_{P+1}, \hat{\upsilon}_{P+1}). \tag{76}$$

By Assumption 1, 2 and Theorem 1, [25], we have

$$\sup_{\omega,\upsilon} I_w(\omega, \upsilon) = O(\log NM). \tag{77}$$

Therefore,

$$H_2(\hat{\rho}_{P+1}, \hat{\omega}_{P+1}, \hat{\upsilon}_{P+1}, \hat{\varphi}_{P+1}) = O\left(\frac{\log NM}{NM}\right) \ . \tag{78}$$

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 $\hat{\rho}_{P+1}, \hat{\omega}_{P+1}, \hat{\upsilon}_{P+1}, \hat{\varphi}_{P+1}$, defined above, *i.e.*,

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

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

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

$$\sum_{n=0}^{N-1} \cos(\omega n + \varphi) = \frac{\sin\left([N - \frac{1}{2}]\omega + \varphi\right) + \sin\left(\frac{\omega}{2} - \varphi\right)}{2\sin\left(\frac{\omega}{2}\right)} = O(1). \tag{79}$$

Hence, as $N \to \infty$

$$\frac{1}{\log N} \sum_{n=0}^{N-1} \cos(\omega n + \varphi) = o(1) \ , \tag{80}$$

and consequently

$$\frac{1}{N} \sum_{n=0}^{N-1} \cos(\omega n + \varphi) = o\left(\frac{\log N}{N}\right) \ . \tag{81}$$

Next, we evaluate $H_3$. Consider the first term in (72). By (81) we have

$$\frac{1}{2NM} \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} \rho_{P+1}^2 \cos(2\omega_{P+1} n + 2\upsilon_{P+1} m + 2\varphi_{P+1}) = o\left(\frac{\log NM}{NM}\right), \tag{82}$$

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

Consider the second term in (72). By (81) and unless there exists some $i$, $1 \leq i \leq P$, such that $(\omega_{P+1}, \upsilon_{P+1}) = (\omega_i^0, \upsilon_i^0)$, we have as $\Psi(N, M) \to \infty$,

$$\frac{1}{NM} \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} \sum_{i=1}^{P} \rho_i^0 \rho_{P+1} \cos(\omega_i^0 n + \upsilon_i^0 m + \varphi_i^0) \cos(\omega_{P+1} n + \upsilon_{P+1} m + \varphi_{P+1}) = o\left(\frac{\log NM}{NM}\right), \tag{83}$$

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

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 no two different regressors with identical spatial frequencies, it follows that one of the estimated frequencies $(\omega_i, \upsilon_i)$ is due to noise contribution. Hence, by interchanging the roles of $(\omega_{P+1}, \upsilon_{P+1})$ and $(\omega_i, \upsilon_i)$, and repeating the above argument we conclude that this term has the same order as in (83). Similarly, for the third term in (72): By (81) 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 $\Psi(N, M) \to \infty$,

$$\frac{1}{NM} \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} \sum_{i=1}^{P} \rho_i \rho_{P+1} \cos(\omega_i n + \upsilon_i m + \varphi_i) \cos(\omega_{P+1} n + \upsilon_{P+1} m + \varphi_{P+1}) = o\left(\frac{\log NM}{NM}\right). \quad (84)$$

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 $\Psi(N, M) \to \infty$

$$H_3(\theta_{P+1}) = o\left(\frac{\log NM}{NM}\right). \quad (85)$$

On the other hand, the strong consistency (73) of the LSE under the correct model order assumption implies that as $\Psi(N, M) \to \infty$ the minimal value of $\mathcal{L}_P(\theta_P) = \sigma^2 \sum_{(r,s)\in D} a^2(r, s)$ a.s., while from (78) we have for the minimal value of $H_2$ that $H_2(\theta_{P+1}) = O\left(\frac{\log NM}{NM}\right)$. Hence, the value of $H_3(\theta_{P+1})$ at *any* point in $\Theta_{p+1}$ 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 (69) as $\Psi(N, M) \to \infty$ we have

$$\begin{aligned} \mathcal{L}_{P+1}(\theta_{P+1}) &= \mathcal{L}_P(\theta_P) + H_2(\rho_{P+1}, \omega_{P+1}, \upsilon_{P+1}, \varphi_{P+1}) + H_3(\theta_{P+1}) \\ &= \mathcal{L}_P(\theta_P) + H_2(\rho_{P+1}, \omega_{P+1}, \upsilon_{P+1}, \varphi_{P+1}) + o\left(\frac{\log NM}{NM}\right). \end{aligned} \quad (86)$$

Since $\mathcal{L}_P(\theta_P)$ is a function of the parameter vector $\theta_P$ and is independent of $\rho_{P+1}, \omega_{P+1}, \upsilon_{P+1}, \varphi_{P+1}$, while $H_2$ is a function of $\rho_{P+1}, \omega_{P+1}, \upsilon_{P+1}, \varphi_{P+1}$ and is independent of $\theta_P$, the problem of minimizing $\mathcal{L}_{P+1}(\theta_{P+1})$ becomes *separable* as $\Psi(N, M) \to \infty$. Thus minimizing (86) is equivalent to separately minimizing $\mathcal{L}_P(\theta_P)$ and $H_2(\rho_{P+1}, \omega_{P+1}, \upsilon_{P+1}, \varphi_{P+1})$ as $\Psi(N, M) \to \infty$. Using the foregoing conclusions, the theorem follows. ■

## Appendix C: Proof of Lemma 3

The next lemma is essential in order to prove Theorem 1 and Theorem 3. It is an extension for the 2-D case of a lemma originally proposed by Hannan, [9] for the case of 1-D signals.

Let $D$ be an *infinite* order non-symmetrical half-plane support defined as in (3) and let $D(k, l)$ be a *finite* order non-symmetrical half-plane support, defined by

$$D(k, l) = \left\{(i, j) \in \mathbb{Z}^2 : i = 0, 0 \leq j \leq l\right\} \cup \left\{(i, j) \in \mathbb{Z}^2 : 0 < i \leq k, -l \leq j \leq l\right\} \quad (87)$$

**Lemma 3.** *Let the field $\{w(n, m)\}$ be defined as in (4), and the field $\{u(n, m)\}$ is an i.i.d. real valued zero-mean random field with finite second order moment, $\sigma^2$. The sequence $a(i, j)$ is a* *square summable deterministic sequence, i.e.*

$$\sum_{(r,s)\in D} a^2(r,s) < \infty. \tag{88}$$

*Then,*

$$\sup_{\omega,\upsilon}\left|\frac{1}{NM}\sum_{n=0}^{N-1}\sum_{m=0}^{M-1} w(n,m)\cos\left(\omega n+\nu m\right)\right| \to 0 \ \textit{a.s. as}\ \Psi(N,M)\to\infty \tag{89}$$

*Proof:*

First, it is easy to see that,

$$\sup_{\omega,\upsilon}\left|\frac{1}{NM}\sum_{n=0}^{N-1}\sum_{m=0}^{M-1} w(n,m)\cos\left(\omega n+\nu m\right)\right| \leq$$
$$\sup_{\omega,\upsilon}\left|\frac{1}{2NM}\sum_{n=0}^{N-1}\sum_{m=0}^{M-1} w(n,m)e^{j(\omega n+\nu m)}\right| + \sup_{\omega,\upsilon}\left|\frac{1}{2NM}\sum_{n=0}^{N-1}\sum_{m=0}^{M-1} w(n,m)e^{-j(\omega n+\nu m)}\right|. \tag{90}$$

Hence it is sufficient to prove the lemma for exponentials, *i.e.*, we wish to prove that

$$\sup_{\omega,\upsilon}\left|\frac{1}{NM}\sum_{n=0}^{N-1}\sum_{m=0}^{M-1} w(n,m)e^{j(\omega n+\nu m)}\right| \to 0 \text{ a.s. as } \Psi(N,M)\to\infty \tag{91}$$

Define the set $D(k,l)^C = D\setminus D(k,l)$. Then,

$$w(n,m) = \sum_{D(k,l)} a(r,s)u(n-r,m-s) + \sum_{D(k,l)^C} a(r,s)u(n-r,m-s) = v(n,m)+z(n,m). \tag{92}$$

Then,

$$\sup_{\omega,\upsilon}\left|\frac{1}{NM}\sum_{n=0}^{N-1}\sum_{m=0}^{M-1} z(n,m)e^{j(\omega n+\nu m)}\right| \leq \frac{1}{NM}\sum_{n=0}^{N-1}\sum_{m=0}^{M-1}|z(n,m)| \leq \left\{\frac{1}{NM}\sum_{n=0}^{N-1}\sum_{m=0}^{M-1} z^2(n,m)\right\}^{\frac{1}{2}}. \tag{93}$$

By the SLLN, the R.H.S. of the last inequality convergence, almost surely, to

$$E[z(0,0)^2]^{\frac{1}{2}} = \left(\sigma^2\sum_{D(k,l)^C} a^2(r,s)\right)^{\frac{1}{2}}, \tag{94}$$

which due to (88) may be made arbitrary small by taking $k$ and $l$ sufficiently large.

Hence it is sufficient to prove the lemma with $w(n,m)$ replaced by $v(n,m)$.

$$\sup_{\omega,\upsilon}\left|\frac{1}{NM}\sum_{n=0}^{N-1}\sum_{m=0}^{M-1}v(n,m)e^{j(\omega n+\nu m)}\right| \leq \sum_{D(k,l)}|a(r,s)|\sup_{\omega,\upsilon}\left|\frac{1}{NM}\sum_{n=0}^{N-1}\sum_{m=0}^{M-1}u(n-r,m-s)e^{j(\omega n+\nu m)}\right|. \quad (95)$$

Since the summation is finite and $\{u(n,m)\}$ is i.i.d., it is sufficient to prove the lemma with $w(n,m)$ replaced by $u(n,m)$. Thus, we consider the mean square of the discussed supremum

$$\begin{aligned} &E\left[\sup_{\omega,\upsilon}\left|\frac{1}{NM}\sum_{n=0}^{N-1}\sum_{m=0}^{M-1}u(n,m)e^{j(\omega n+\nu m)}\right|^2\right] \\ &= E\left[\sup_{\omega,\upsilon}\frac{1}{(NM)^2}\sum_{n=0}^{N-1}\sum_{m=0}^{M-1}\sum_{k=0}^{N-1}\sum_{l=0}^{M-1}u(n,m)u(k,l)e^{j(\omega(n-k)+\nu(m-l))}\right]. \end{aligned} \quad (96)$$

By letting,

$$\begin{aligned} n-k&=p, \\ m-l&=r, \end{aligned} \quad (97)$$

we change the order of summation to that

$$\begin{aligned} \sum_{n=0}^{N-1}\sum_{k=0}^{N-1} &= \sum_{|p|<N}\sum_{n\in S_N}, \\ \sum_{m=0}^{M-1}\sum_{l=0}^{M-1} &= \sum_{|r|<M}\sum_{m\in S_M}, \end{aligned} \quad (98)$$

where,

$$\begin{aligned} S_N &= \{n\in\mathbb{Z} : \max(0,p)\leq n\leq \min(N-1,p+N-1)\}, \\ S_M &= \{m\in\mathbb{Z} : \max(0,r)\leq m\leq \min(M-1,r+M-1)\}. \end{aligned} \quad (99)$$

Therefore

$$\begin{aligned} \sum_{n\in S_N}1 &= \begin{cases} N-p, & p\geq 0 \\ N+p, & p<0 \end{cases} = N-|p|, \\ \sum_{m\in S_M}1 &= \begin{cases} M-r, & r\geq 0 \\ M+r, & r<0 \end{cases} = M-|r|. \end{aligned} \quad (100)$$

Rewriting (96) we have

$$\frac{1}{(NM)^2}E\left[\sup_{\omega,v}\sum_{|p|<N}\sum_{|r|<M}\sum_{n\in S_N}\sum_{m\in S_M}u(n,m)u(n-p,m-r)e^{j(\omega p+\nu r)}\right]$$

$$=\frac{1}{(NM)^2}E\left[\sum_{n=0}^{N-1}\sum_{m=0}^{M-1}u(n,m)^2+\sup_{\omega,v}\sum_{\substack{|p|<N\\p\neq 0}}\sum_{\substack{|r|<M\\r\neq 0}}\sum_{n\in S_N}\sum_{m\in S_M}u(n,m)u(n-p,m-r)e^{j(\omega p+\nu r)}\right]$$

$$\leq\frac{1}{(NM)^2}\left\{NM\sigma^2+\sum_{\substack{|p|<N\\p\neq 0}}\sum_{\substack{|r|<M\\r\neq 0}}E\left[\left|\sum_{n\in S_N}\sum_{m\in S_M}u(n,m)u(n-p,m-r)\right|\right]\right\}, \tag{101}$$

where we have employed the triangle inequality.

Let investigate the second term on the R.H.S. of (101). From the Cauchy-Schwartz inequality, for any r.v. $x$, $E\left[|x|\right]\leq E\left[|x|^2\right]^{\frac{1}{2}}$, hence

$$E\left[\left|\sum_{n\in S_N}\sum_{m\in S_M}u(n,m)u(n-p,m-r)\right|\right]\leq E\left[\left|\sum_{n\in S_N}\sum_{m\in S_M}u(n,m)u(n-p,m-r)\right|^2\right]^{\frac{1}{2}}$$

$$=\left(\sum_{n\in S_N}\sum_{m\in S_M}\sum_{n'\in S_N}\sum_{m'\in S_M}E[u(n,m)u(n-p,m-r)u(n',m')u(n'-p,m'-r)]\right)^{\frac{1}{2}}$$

$$=\left(\sum_{n\in S_N}\sum_{m\in S_M}\sigma^4\right)^{\frac{1}{2}}=\sigma^2(N-|p|)^{\frac{1}{2}}(M-|r|)^{\frac{1}{2}}. \tag{102}$$

which follows from the observation that for $p,r\neq 0$, the fourth order moment of the field $\{u(n,m)\}$ equals zero for all $n\neq n'$ or $m\neq m'$.

Hence we can finally write

$$E\left[\sup_{\omega,v}\left|\frac{1}{NM}\sum_{n=0}^{N-1}\sum_{m=0}^{M-1}u(n,m)e^{j(\omega n+\nu m)}\right|^2\right]$$

$$\leq\frac{1}{(NM)^2}\left\{NM\sigma^2+\sum_{\substack{|p|<N\\p\neq 0}}\sum_{\substack{|r|<M\\r\neq 0}}\sigma^2(N-|p|)^{\frac{1}{2}}(M-|r|)^{\frac{1}{2}}\right\}$$

$$\leq\frac{\sigma^2}{(NM)^2}\{NM+4(NM)^{\frac{3}{2}}\}\leq\frac{K}{(NM)^{\frac{1}{2}}}=O(N^{-\frac{1}{2}}M^{-\frac{1}{2}}). \tag{103}$$

where $K$ some finite positive constant.

Now following the ideas of Doob, [4]( ch. X, 6), let $R$ and $S$ be some positive integers such that $N > R^\delta$, and $M > S^\delta$, for $\delta > 2$. Hence, for any such choice of $N$ and $M$, from (103),

$$E\left[\sup_{\omega,\upsilon}\left|\frac{1}{NM}\sum_{n=0}^{N-1}\sum_{m=0}^{M-1}u(n,m)e^{j(\omega n+\nu m)}\right|^2\right] \leq \frac{K}{(RS)^{\frac{\delta}{2}}}. \tag{104}$$

Hence, if we take $N = N(R)$ and $M = M(S)$ to be the smallest integers not smaller then $R^\delta$ and $S^\delta$, respectively, then (104) still holds.

Hence, by Chebyshev inequality for every $\epsilon > 0$

$$\begin{aligned} & P\left(\sup_{\omega,\upsilon}\left|\frac{1}{N(R)M(S)}\sum_{n=0}^{N(R)-1}\sum_{m=0}^{M(S)-1}u(n,m)e^{j(\omega n+\nu m)}\right| \geq \epsilon\right) \\ & \leq \frac{E\left[\sup_{\omega,\upsilon}\left|\frac{1}{N(R)M(S)}\sum_{n=0}^{N(R)-1}\sum_{m=0}^{M(S)-1}u(n,m)e^{j(\omega n+\nu m)}\right|^2\right]}{\epsilon^2} \leq \frac{K}{\epsilon^2(RS)^{\frac{\delta}{2}}} \end{aligned} \tag{105}$$

and then since $\delta > 2$

$$\sum_{R=1}^{\infty}\sum_{S=1}^{\infty}P\left(\sup_{\omega,\upsilon}\left|\frac{1}{N(R)M(S)}\sum_{n=0}^{N(R)-1}\sum_{m=0}^{M(S)-1}u(n,m)e^{j(\omega n+\nu m)}\right| > \epsilon\right) \leq \sum_{R=1}^{\infty}\sum_{S=1}^{\infty}\frac{K}{\epsilon^2(RS)^{\frac{\delta}{2}}} < \infty. \tag{106}$$

Hence, by the Borel-Cantelli lemma,

$$\sup_{\omega,\upsilon}\left|\frac{1}{N(R)M(S)}\sum_{n=0}^{N(R)-1}\sum_{m=0}^{M(S)-1}u(n,m)e^{j(\omega n+\nu m)}\right| \to 0 \text{ a.s. as } \Psi(R,S) \to \infty. \tag{107}$$

Now,

$$\sup_{\substack{N(R)\le N\le N(R+1)\\ M(S)\le M\le M(S+1)}} \sup_{\omega,\upsilon} \left| \frac{1}{NM}\sum_{n=0}^{N-1}\sum_{m=0}^{M-1} u(n,m)e^{j(\omega n+\nu m)} - \frac{1}{NM}\sum_{n=0}^{N(R)-1}\sum_{m=0}^{M(S)-1} u(n,m)e^{j(\omega n+\nu m)} \right|$$

$$\le \sup_{\substack{N(R)\le N\le N(R+1)\\ M(S)\le M\le M(S+1)}} \sup_{\omega,\upsilon} \frac{1}{NM} \left| \sum_{n=0}^{N(R)-1}\sum_{m=M(S)}^{M-1} u(n,m)e^{j(\omega n+\nu m)} \right|$$

$$+ \sup_{\substack{N(R)\le N\le N(R+1)\\ M(S)\le M\le M(S+1)}} \sup_{\omega,\upsilon} \frac{1}{NM} \left| \sum_{n=N(R)}^{N-1}\sum_{m=0}^{M(S)-1} u(n,m)e^{j(\omega n+\nu m)} \right|$$

$$+ \sup_{\substack{N(R)\le N\le N(R+1)\\ M(S)\le M\le M(S+1)}} \sup_{\omega,\upsilon} \frac{1}{NM} \left| \sum_{n=N(R)}^{N-1}\sum_{m=M(S)}^{M-1} u(n,m)e^{j(\omega n+\nu m)} \right| = I_1 + I_2 + I_3. \tag{108}$$

Consider the first term in the previous equation. Using the triangle inequality

$$I_1 \le \frac{1}{M(S)} \sum_{m=M(S)}^{M(S+1)-1} \left( \sup_{\omega} \frac{1}{N(R)} \left| \sum_{n=0}^{N(R)-1} u(n,m)e^{j\omega n} \right| \right). \tag{109}$$

Let

$$\tilde{u}(m) = \sup_{\omega} \frac{1}{N(R)} \left| \sum_{n=0}^{N(R)-1} u(n,m)e^{j\omega n} \right|. \tag{110}$$

Since $\{u(n,m)\}$ is i.i.d., it is clear that $\{\tilde{u}(m)\}$ is an i.i.d. sequence of random variables. Moreover, from [9] (or by repeating the derivation in (92)-(104) for the process $u(n,m)$ with a fixed $m$) we have

$$E\left[\tilde{u}(m)^2\right] = E\left[ \sup_{\omega} \frac{1}{N(R)} \left| \sum_{n=0}^{N(R)-1} u(n,m)e^{j\omega n} \right|^2 \right] \le \frac{K_1}{R^{\frac{\delta}{2}}}. \tag{111}$$

Taking the mean of the square of $I_1$ we have

$$\begin{aligned}
E\left[|I_1|^2\right] &\le \frac{1}{M(S)^2} \sum_{m=M(S)}^{M(S+1)-1} \sum_{m'=M(S)}^{M(S+1)-1} E\left[\tilde{u}(m)\tilde{u}(m')\right] \\
&\le \frac{1}{M(S)^2} \sum_{m=M(S)}^{M(S+1)-1} \sum_{m'=M(S)}^{M(S+1)-1} E\left[\tilde{u}(m)^2\right]^{\frac{1}{2}} E\left[\tilde{u}(m')^2\right]^{\frac{1}{2}} \\
&\le \frac{K_1(M(S+1)-1-M(S))^2}{R^{\frac{\delta}{2}} M(S)^2} \le \frac{K}{R^{\frac{\delta}{2}} S^2}.
\end{aligned} \tag{112}$$

Using once again the Chebyshev inequality and the Borel-Cantelli lemma we have that $I_1 \to 0$ a.s. as $\Psi(R,S) \to \infty$. Repeating the same consideration for $I_2$ we have that $I_2 \to 0$ a.s. as $\Psi(R,S) \to \infty$. Finally, for $I_3$ we have

$$\begin{aligned}
E[|I_3|^2] &\leq E\left[\left|\frac{1}{N(R)M(S)} \sum_{n=N(R)}^{N(R+1)-1} \sum_{m=M(S)}^{M(R+1)-1} |u(n,m)|\right|^2\right] \\
&= \frac{1}{(N(R)M(S))^2} \sum_{n=N(R)}^{N(R+1)-1} \sum_{m=M(S)}^{M(S+1)-1} \sum_{n'=N(R)}^{N(R+1)-1} \sum_{m'=M(S)}^{M(S+1)-1} E[|u(n,m)u(n',m')|] \\
&\leq \sigma^2 \frac{(N(R+1)-1-N(R))^2(M(S+1)-1-M(S))^2}{(N(R)M(S))^2} \leq \frac{K}{(RS)^2}.
\end{aligned} \tag{113}$$

Using again the Chebyshev inequality and the Borel-Cantelli lemma we have that $I_3 \to 0$ a.s. as $\Psi(R,S) \to \infty$.

Finally, we have that

$$\sup_{\omega,\upsilon} \left| \frac{1}{NM} \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} u(n,m)e^{j(\omega n+\nu m)} - \frac{1}{NM} \sum_{n=0}^{N(R)-1} \sum_{m=0}^{M(S)-1} u(n,m)e^{j(\omega n+\nu m)} \right| \to 0 \text{ a.s.} \tag{114}$$

for all $N(R) \leq N < N(R+1)$ and $M(S) \leq M < M(S+1)$, as $\Psi(R,S) \to \infty$, and hence as $\Psi(N,M) \to \infty$.

Since $\frac{N(R)}{N(R+1)} \to 1$ and $\frac{M(S)}{M(S+1)} \to 1$ as $\Psi(R,S) \to \infty$ we can replace $\frac{1}{NM}$ in the second term by $\frac{1}{N(R)M(S)}$. Therefore, we have

$$\sup_{\omega,\upsilon} \left| \frac{1}{NM} \sum_{n=0}^{N-1} \sum_{m=0}^{M-1} u(n,m)e^{j(\omega n+\nu m)} - \frac{1}{N(R)M(S)} \sum_{n=0}^{N(R)-1} \sum_{m=0}^{M(S)-1} u(n,m)e^{j(\omega n+\nu m)} \right| \to 0 \text{ a.s.} \tag{115}$$

From (115) and (107) the lemma follows.

■

# References


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

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

[3] G. Cohen and J. M. Francos, "Least Squares Estimation of 2-D Sinusoids in Colored Noise: Asymptotic Analysis," *IEEE Trans. Info. Theory*, vol. 48, pp. 2243-2252, August 2002.

[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] D. Kundu and S. Nandi, *Statistical Signal Processing: Frequency Estimation*, Springer, 2012.

[17] 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.

[18] 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.

[19] B. Nadler and L. Kontorovich, “Model Selection for Sinusoids in Noise: Statistical Analysis and a New Penalty Term,” *IEEE Trans. Signal Process.*, vol. 59, pp. 1333-1345, 2011.

[20] A. Prasad, D. Kundu and A. Mitra, “ Sequential estimation of two dimensional sinusoidal models,” *J. of Probability and Statistical Sciences*, to appear.

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

[22] 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.

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

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

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

[26] 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.

[27] 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.

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

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