STRONG CONSISTENCY OF A FAMILY OF MODEL ORDER SELECTION RULES FOR ESTIMATING THE PARAMETERS OF 2-D SINUSOIDS IN WHITE NOISE

Mark Kliger and Joseph M. Francos

Electrical & Computer Engineering Dept.
Ben-Gurion University
Beer Sheva 84105, Israel

ABSTRACT

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 white Gaussian noise field. In this paper we prove the strong consistency of a large family of model order selection rules, which includes the MAP based rule as a special case.

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 white Gaussian noise field. This problem is, in fact, a special case of a much more general problem: From the 2-D Wold-like decomposition we have that any 2-D regular and homogeneous discrete random field 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 assumed to be a white noise field.

Many algorithms have been devised to estimate the parameters of sinusoids observed in additive white Gaussian noise. Most of these assume the number of sinusoids is *a-priori* known. However this assumption does not always hold in practice. In the past three 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), the minimum description length (MDL), and the maximum a-posteriori probability criterion (MAP). All these criteria have a common form composed of two terms: a data term and a penalty term, where the data term is the log-likelihood function evaluated for the assumed model.

Most of the papers that address the problem of model order selection are concerned with various models of one-dimensional signals, while the problem of modelling multidimensional fields has received considerably less attention. Stoica *et al.*, [9] proposed a cross-validation selection rule and demonstrated its asymptotic equivalence to the Generalized Akaike Information Criterion (GAIC). The suggested criterion is not derived for any specific model. The penalty term is given by $kK(N)$ where k is the number of model parameters, N is the length of the observed data vector, and $K(N)$ is some penalty term which is a function of N. In [5] this criterion is employed to detect the number of sinusoids in 1-D and 2-D signals. The penalty term for 2-D signals is the same as in the 1-D case. The penalty parameter is chosen as $K(N) = c \log\log N$ where $c > 2$. Stoica *et al.* in [9] and Li *et al.* in [5] arrived at this choice of $K(N)$, by using consistency arguments based on [1]. However, in [1] consistency of an order selection criterion for ARMA models is proved, while the model considered in [5] is that of sinusoids in noise. Moreover, for the data model of 1-D sinusoids observed in white noise, Quinn, [6], derives conditions for strong consistency of any model order selection criterion. The penalty term of the criterion in [5], does not satisfy Quinn's consistency conditions even for the 1-D problem. In [3], a maximum *a-posteriori* (MAP) model order selection criterion for jointly estimating the number and the parameters of two-dimensional sinusoids observed in the presence of an additive white Gaussian noise field, is derived.

In this paper, we establish the strong consistency of a large family of model order selection rules, which includes the MAP based rule as a special case.

2. NOTATIONS AND DEFINITIONS

Let $\{y(s,t)\}$, $(s,t) \in \Psi(S,T)$ where $\Psi(S,T) = \{(i,j)|0 \le i \le S-1, 0 \le j \le T-1\}$ be the observed 2-D real valued random field such that

$$y(s,t) = h(s,t) + u(s,t). \tag{1}$$

The field $\{u(s,t)\}$ is a 2-D zero mean, white Gaussian field with finite variance σ^2. The field $\{h(s,t)\}$ is the harmonic random field. Assuming there are k sinusoidal components in the harmonic field we have

$$h(s,t) = \sum_{i=1}^{k} C_i \cos(s\omega_i + t\nu_i) + G_i \sin(s\omega_i + t\nu_i), \tag{2}$$

where (ω_i, ν_i) are the spatial frequencies of the ith harmonic. The C_i's and G_i's are the unknown amplitudes of the sinusoidal components in the observed realization.

Let us define the following matrix notations:

$$\begin{aligned}\mathbf{y} &= [y(0,0) \dots y(0,T-1)\; y(1,0) \dots y(1,T-1)\; \dots \\ &\quad \dots y(S-1,T-1)]^T. \end{aligned} \tag{3}$$

The vectors $\mathbf{u}$ and $\mathbf{h}$ are similarly defined. Rewriting (1) we have $\mathbf{y} = \mathbf{h} + \mathbf{u}$. Also define

$$\mathbf{a_k} = [C_1\; G_1\; C_2\; G_2\; \dots\; C_k\; G_k]^T\;. \tag{4}$$

Let

$$\begin{aligned}\mathbf{e}_i &= [e^{j[0\omega_i + 0\nu_i]}\; e^{j[0\omega_i + 1\nu_i]} \dots e^{j[0\omega_i + (T-1)\nu_i]} \dots \\ &\quad \dots e^{j[(S-1)\omega_i + (T-1)\nu_i]}]^T, \end{aligned} \tag{5}$$

and let us define the following $ST \times 2k$ matrix

$$\mathbf{D}_k = [\mathrm{Re}(\mathbf{e}_1)\ \mathrm{Im}(\mathbf{e}_1)\ \mathrm{Re}(\mathbf{e}_2)\ \mathrm{Im}(\mathbf{e}_2)\ \ldots\ \mathrm{Re}(\mathbf{e}_k)\ \mathrm{Im}(\mathbf{e}_k)]\,. \tag{6}$$

Using the foregoing notations we have that

$$\mathbf{y} = \mathbf{D}_k \mathbf{a}_k + \mathbf{u}. \tag{7}$$

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

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

Definition 1: The sequence of subsets $\{\Psi_i\}$ is said to tend to infinity (we adopt the notation $\Psi_i \to \infty$) as $i \to \infty$ if

$$\lim_{i\to\infty} \min(S_i, T_i) = \infty$$

and

$$0 < \lim_{i\to\infty} (S_i/T_i) < \infty.$$

To simplify notations, we shall omit in the following the subscript i. Thus, the notation $\Psi(S,T) \to \infty$ implies that both S and T tend to infinity as functions of i, and at roughly the same rate.

Let $\boldsymbol{\theta}_k \in \boldsymbol{\Theta}_k$ denote the parameter vector of the harmonic field, *i.e.*,

$$\boldsymbol{\theta}_k = [C_1\ G_1\ \omega_1\ \nu_1\ \ldots\ C_k\ G_k\ \omega_k\ \nu_k]^T, \tag{8}$$

where for all l, C_l, G_l are real and bounded. Assume further that $\omega_l, \nu_l \in (0, 2\pi)$ where $\min(|\omega_l - \omega_j|) \ge \delta$ or $\min(|\nu_l - \nu_j|) \ge \delta$ for $l \ne j$. Hence, the parameter space, $\boldsymbol{\Theta}_k$, is a subset of the $4k$ dimensional Euclidian space. By the above assumption we further conclude that $\mathbf{D}_k$ has rank $2k$, and that the corresponding $2k \times 2k$ Gram matrix $\mathbf{D}_k^T\mathbf{D}_k$ is of rank $2k$ as well.

Let m denote the actual number of sinusoidal components in the observed field and let $A_i = \sqrt{C_i^2 + G_i^2}$ denote the amplitude of the ith sinusoid. It is assumed that the amplitudes are strictly positive and bounded. For convenience, and without loss of generality, it is further assumed that the sinusoidal components are indexed according to a descending order of their amplitudes where $A_1 \ge A_2 \ge \ldots \ge A_m > 0$.

Let $\mathbf{P}_k^\perp$ denote the projection matrix defined by

$$\mathbf{P}_k^\perp = \mathbf{I}_{2k} - \mathbf{D}_k(\mathbf{D}_k^T\mathbf{D}_k)^{-1}\mathbf{D}_k^T. \tag{9}$$

where $\mathbf{I}_{2k}$ is an $2k \times 2k$ identity matrix. Under the normality assumptions of the noise field, the maximum likelihood estimate (MLE) $\hat{\boldsymbol{\theta}}_k$ of $\boldsymbol{\theta}_k$ is the same as the nonlinear least square estimate (LSE) and is obtained through minimization of the quadratic form $\mathbf{y}^T\mathbf{P}_k^\perp\mathbf{y}$. Let $\hat{\mathbf{P}}_k^\perp$ denote the matrix $\mathbf{P}_k^\perp$, with $\boldsymbol{\theta}_k$ substituted by $\hat{\boldsymbol{\theta}}_k$.

Let $p(k)$ denote the *a-priori* probability of the kth model, where k denotes the unknown number of sinusoidal components in the data model given by (1), (2).

It is assumed that there are Q competing models, where $Q > m$, and that each model is equiprobable. That is

$$p(k) = \frac{1}{Q}, \quad k \in \mathcal{Z}_Q, \tag{10}$$

where

$$\mathcal{Z}_Q = \{0, 1, 2, \ldots, Q-1\}.$$

Following the MDL-MAP template, define the statistic

$$\chi_\xi(k) = ST\log(\mathbf{y}^T\hat{\mathbf{P}}_k^\perp\mathbf{y}) + \xi k \log ST, \tag{11}$$

where $\xi > 8$ is a finite constant and $k \in \mathcal{Z}_Q$.

The number of 2-D sinusoids m is estimated by minimizing of $\chi_\xi(k)$ over $k \in \mathcal{Z}_Q$

$$\begin{aligned}\hat{m} &= \arg\min_{k\in\mathcal{Z}_Q}\left\{\chi_\xi(k)\right\}\\ &= \arg\min_{k\in\mathcal{Z}_Q}\left\{ST\log(\mathbf{y}^T\hat{\mathbf{P}}_k^\perp\mathbf{y}) + \xi k\log ST\right\}.\end{aligned} \tag{12}$$

We finally note that the choice of $\xi = 10$ in (12) yields the MAP model order selection rule derived in [3], while the choice of $\xi = 4$ in (12) corresponds to a "naive" application of a model order selection rule based on the MDL principle, [8].

3. CONSISTENCY OF THE CRITERION

The objective of this section is to prove the asymptotic consistency of the model order selection procedure in (12).

Theorem 1 *Let m denote the correct number of sinusoids in the field, and let $\hat{m}$ be given by (12) with $\xi > 8$. Then as $\Psi(S,T) \to \infty$*

$$\hat{m} \to m \ \ a.s. \tag{13}$$

Proof: Let

$$\hat{\sigma}_k^2 = \frac{\mathbf{y}^T\hat{\mathbf{P}}_k^\perp\mathbf{y}}{ST} \tag{14}$$

denote the variance of the residual field obtained after the removal of the k most dominant estimated sinusoidal components of the observed field.

Let $\hat{A}_i$ denote the ML estimate of A_i. Hence, for $k \le m$ we have as $\Psi(S,T) \to \infty$

$$\hat{\sigma}_k^2 = \sigma^2 + \frac{1}{2}\sum_{i=1}^{m} A_i^2 - \frac{1}{2}\sum_{i=1}^{k}\hat{A}_i^2 + o(1) \ \text{ a.s.} \tag{15}$$

(See Appendix A for the detailed derivation of (15)). Since under the normality assumption of the noise field, the maximum likelihood estimate $\hat{\boldsymbol{\theta}}_k$ of $\boldsymbol{\theta}_k$ is the same as the nonlinear least square estimate, we conclude using Theorem 1, [4] that as $\Psi(S,T) \to \infty$

$$\hat{A}_i \to A_i \ \text{ a.s. } \ i = 1\ldots k \tag{16}$$

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

$$\hat{\sigma}_k^2 \to \sigma^2 + \frac{1}{2}\sum_{i=k+1}^{m} A_i^2 \ \text{ a.s.} \tag{17}$$

and similarly

$$\hat{\sigma}_{k-1}^2 \to \sigma^2 + \frac{1}{2}\sum_{i=k}^{m} A_i^2 \ \text{ a.s.} \tag{18}$$

Using (14) we can write the criterion function in the next form

$$\begin{aligned}\chi_\xi(k) &= ST\log(ST\hat{\sigma}_k^2) + \xi k\log ST\\ &= ST\log ST + ST\log\hat{\sigma}_k^2 + \xi k\log ST.\end{aligned} \tag{19}$$

Therefore, for $k \le m$,

$$\begin{aligned}&\chi_\xi(k-1)-\chi_\xi(k)\\&= ST\log ST + ST\log\hat\sigma^2_{k-1}+\xi(k-1)\log ST\\&-ST\log ST - ST\log\hat\sigma^2_k - \xi k\log ST\\&= ST\log\left(\frac{\hat\sigma^2_{k-1}}{\hat\sigma^2_k}\right)-\xi\log ST.\end{aligned}\tag{20}$$

Since $\frac{\log ST}{ST}$ tends to zero, as $\Psi(S,T)\to\infty$, then as $\Psi(S,T)\to\infty$

$$(ST)^{-1}(\chi_\xi(k-1)-\chi_\xi(k)) \to \log\left(1+\frac{A_k^2}{2\sigma^2+\sum_{i=k+1}^m A_i^2}\right)\ \text{a.s.}\tag{21}$$

Since $\log\left(1+\frac{A_k^2}{2\sigma^2+\sum_{i=k+1}^m A_i^2}\right)$ is strictly positive, we have $\chi_\xi(k-1)>\chi_\xi(k)$. Hence, for $k\le m$, the function $\chi_\xi(k)$ is monotonically decreasing with k.

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

Let

$$I_u(\omega,\nu)=\frac{2}{ST}\left|\sum_{s=0}^{S-1}\sum_{t=0}^{T-1}u(s,t)e^{-j[s\omega+t\nu]}\right|^2\tag{22}$$

denote the periodogram of the noise field (scaled by a factor of 2). Based on [2], Theorem 2.2 we have that if $\{u(s,t)\}$ is a two-dimensional *i.i.d.* zero-mean real valued field with variance σ^2 such that

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

then

$$\limsup_{\Psi(S,T)\to\infty}\frac{\sup_{\omega,\nu} I_u(\omega,\nu)}{\sigma^2\log(ST)}\le 8\ \text{a.s.}\tag{23}$$

It is easy to check that indeed when $\{u(s,t)\}$ is a Gaussian white noise field the condition $E[u(0,0)^2\log|u(0,0)|]<\infty$ is satisfied. Indeed,

$$E[u^2\log|u|]=E[u^2\log|u|\mathbf{1}_{\{|u|\ge1\}}]+E[u^2\log|u|\mathbf{1}_{\{|u|<1\}}],$$

where

$$E[u^2\log|u|\mathbf{1}_{\{|u|\ge1\}}]\le E[u^4\mathbf{1}_{\{|u|\ge1\}}]\le E[u^4]=3\sigma^4,$$

while for $|u|<1$ the function $u^2\log|u|$ is bounded and has a minimum at $|u|=\exp^{-\frac12}$ and this minimum is $-0.5e^{-1}$. Thus,

$$0\ge E[u^2\log|u|\mathbf{1}_{\{|u|<1\}}]\ge -0.5e^{-1}P(|u|<1\})\ge-\infty.$$

Since the noise field is Gaussian, the MLE $\hat{\boldsymbol\theta}_k$ of $\boldsymbol\theta_k$ and the LSE of $\boldsymbol\theta_k$ are identical. Hence, we have that a.s. as $\Psi(S,T)\to\infty$

$$\hat\sigma^2_{m+l}=\hat\sigma^2_m-\frac{U_l}{ST}+o\left(\frac{\log ST}{ST}\right),\tag{24}$$

where

$$U_l=\sum_{i=1}^{l}I_u(\omega_i,\nu_i)\tag{25}$$

is the sum of the l largest elements of the periodogram of the noise field $\{u(s,t)\}$ (see Appendix B for a detailed derivation of (24)). Clearly

$$U_l\le l\sup_{\omega,\nu}I_u(\omega,\nu).\tag{26}$$

From Theorem 2, [7] as $\Psi(S,T)\to\infty$

$$\hat\sigma^2_m\to\sigma^2\ \text{a.s.}\tag{27}$$

Similarly to (19) and (20), a.s. as $\Psi(S,T)\to\infty$,

$$\begin{aligned}&\chi_\xi(m+l)-\chi_\xi(m)\\&=ST\log ST+ST\log\hat\sigma^2_{m+l}+\xi(m+l)\log ST\\&-ST\log ST-ST\log\hat\sigma^2_m-\xi m\log ST\\&=\xi l\log ST+ST\log\left(1-\frac{U_l}{ST\sigma^2}+o\left(\frac{\log ST}{ST}\right)\right)\\&=\xi l\log ST-\left(\frac{U_l}{\sigma^2}+o(\log ST)\right)(1+o(1))\\&=\log ST\left(\xi l-\frac{U_l}{\sigma^2\log ST}+o(1)\right)\\&\ge\log ST\left(\xi l-\frac{l\sup_{\omega,\nu}I_u(\omega,\nu)}{\sigma^2\log ST}+o(1)\right)\\&=l\log ST\left(\xi-\frac{\sup_{\omega,\nu}I_u(\omega,\nu)}{\sigma^2\log ST}+o(1)\right),\end{aligned}\tag{28}$$

where the second equality is obtained by substituting $\hat\sigma^2_{m+l}$ and $\hat\sigma^2_m$ using the equalities (24) and (27), respectively. The third equality is due to the property that for $x\to0$, $\log(1+x)=x(1+o(1))$, where the observation that the term $\frac{U_l}{\sigma^2 ST}$ tends to zero a.s. as $\Psi(S,T)\to\infty$ is due to (23).

Substituting (23) into (28) we conclude that

$$\chi_\xi(m+l)-\chi_\xi(m)>0\tag{29}$$

for any integer $l\ge1$. Therefore, a.s. as $\Psi(S,T)\to\infty$, the function $\chi_\xi(k)$ has a **global minimum** for $k=m$. ■

4. CONCLUSIONS

In this paper we have established the strong consistency of a family of model order selection rules for the problem of estimating two-dimensional sinusoidal signals, observed in the presence of an additive white Gaussian noise field. The MAP model order selection rule derived in [3] belongs to this family of model order selection criteria.

5. APPENDIX A

$$\begin{aligned}\hat\sigma^2_k &= \frac{\mathbf{y}^T\mathbf{y}}{ST}-\frac{\mathbf{y}^T\hat{\mathbf{P}}_k\mathbf{y}}{ST}=\frac{\mathbf{u}^T\mathbf{u}}{ST}+\frac{\mathbf{u}^T\mathbf{D}_m\mathbf{a}_m}{ST}\\&+\frac{\mathbf{a}_m^T\mathbf{D}_m^T\mathbf{u}}{ST}+\frac{\mathbf{a}_m^T\mathbf{D}_m^T\mathbf{D}_m\mathbf{a}_m}{ST}-\frac{\hat{\mathbf{a}}_k^T\hat{\mathbf{D}}_k^T\hat{\mathbf{D}}_k\hat{\mathbf{a}}_k}{ST},\end{aligned}\tag{30}$$

where

$$\hat{\mathbf{a}}_k=(\hat{\mathbf{D}}_k^T\hat{\mathbf{D}}_k)^{-1}\hat{\mathbf{D}}_k^T\mathbf{y}.\tag{31}$$

By the SLLN , a.s. as $\Psi(S,T)\to\infty$,

$$\frac{\mathbf{u}^T\mathbf{u}}{ST}\to\sigma^2.\tag{32}$$

From Lemma 3, [7] we have that a.s. as $\Psi(S,T)\to\infty$

$$\frac{\mathbf{u}^T\mathbf{D}_m\mathbf{a}_m}{ST}=o(1).\tag{33}$$

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

$$\sum_{s=0}^{S-1} \exp(j\omega s) = O(1). \tag{34}$$

Therefore, for each $\boldsymbol{\theta}_m \in \boldsymbol{\Theta}_m$, as $\Psi(S,T) \to \infty$,

$$\frac{[C_i\ G_i]\,[\mathrm{Re}(\mathbf{e}_i)\ \mathrm{Im}(\mathbf{e}_i)]^T\,[\mathrm{Re}(\mathbf{e}_j)\ \mathrm{Im}(\mathbf{e}_j)]\,[C_i\ G_i]^T}{ST}$$
$$= \begin{cases} \frac{A_i^2}{2} + o(1), & i = j \\ o(1), & i \neq j \end{cases} \tag{35}$$

and then

$$\frac{\mathbf{a}_m^T \mathbf{D}_m^T \mathbf{D}_m^T \mathbf{a}_m}{ST} = \sum_{i=1}^{m} \frac{A_i^2}{2} + o(1). \tag{36}$$

Similarly, since $\hat{\boldsymbol{\theta}}_k$ is the maximum likelihood estimate of $\boldsymbol{\theta}_k$ and $\boldsymbol{\theta}_k \in \boldsymbol{\Theta}_k$

$$\frac{\hat{\mathbf{a}}_k^T \hat{\mathbf{D}}_k^T \hat{\mathbf{D}}^T \hat{\mathbf{a}}_k}{ST} = \sum_{i=1}^{k} \frac{\hat{A}_i^2}{2} + o(1) \ \text{ a.s.} \tag{37}$$

Substituting (33), (36) and (37) into (30) we have (15).

6. APPENDIX B

Let $(\hat{\omega}_{m+1}, \hat{\upsilon}_{m+1})$ denote the largest element of the periodogram of the noise field $\{u(n,m)\}$, *i.e.*,

$$(\hat{\omega}_{m+1}, \hat{\upsilon}_{m+1}) = \underset{(\omega,\upsilon)\in(0,2\pi)^2}{\arg\max}\ I_u(\omega, \upsilon) \tag{38}$$

and let

$$\hat{A}_{m+1} = \frac{2}{ST} \sum_{s=0}^{S-1} \sum_{t=0}^{T-1} u(s,t) e^{-j(s\hat{\omega}_{m+1} + t\hat{\upsilon}_{m+1})}. \tag{39}$$

Let $\hat{C}_{m+1} = \Re(\hat{A}_{m+1})$ and $\hat{G}_{m+1} = \Im(\hat{A}_{m+1})$. Let $\hat{\mathbf{e}}_i$ denote the vector $\mathbf{e}_i$ defined in (5), with (ω_i, ν_i) substituted by $(\hat{\omega}_i, \hat{\nu}_i)$.

Using a straightforward extension of [4], Theorem 2, we have that in the case where the model order is over-estimated the ML estimate contains a subvector that converges a.s. as $\Psi(S,T) \to \infty$ to the correct parameters of the sinusoidal signals, while the frequencies of the sinusoids that result from the over-estimated order assumption are assigned to the spatial frequencies that maximize the noise periodogram. Hence,

$$\hat{\sigma}_{m+1}^2 = \frac{\mathbf{y}^T\mathbf{y}}{ST} - \frac{\hat{\mathbf{a}}_{m+1}^T \hat{\mathbf{D}}_{m+1}^T \hat{\mathbf{D}}_{m+1} \hat{\mathbf{a}}_{m+1}}{ST}$$
$$= \frac{\mathbf{y}^T\mathbf{y}}{ST} - \frac{\hat{\mathbf{a}}_m^T \hat{\mathbf{D}}_m^T \hat{\mathbf{D}}_m \hat{\mathbf{a}}_m}{ST} - M_1 - M_2 = \hat{\sigma}_m^2 - M_1 - M_2, \tag{40}$$

where

$$M_1 = 2\frac{[\hat{C}_{m+1}\ \hat{G}_{m+1}]\,[\mathrm{Re}(\hat{\mathbf{e}}_{m+1})\ \mathrm{Im}(\hat{\mathbf{e}}_{m+1})]^T\,\hat{\mathbf{D}}_m \hat{\mathbf{a}}_m}{ST} \tag{41}$$

and

$$M_2 = \frac{|\hat{A}_{m+1}|^2}{2} + \frac{\hat{C}_{m+1}\hat{G}_{m+1}}{ST} \sum_{s=0}^{S-1} \sum_{t=0}^{T-1} \sin(2s\hat{\omega}_{m+1} + 2t\hat{\upsilon}_{m+1})$$
$$+ \frac{\hat{C}_{m+1}^2 - \hat{G}_{m+1}^2}{2ST} \sum_{s=0}^{S-1} \sum_{t=0}^{T-1} \cos(2s\hat{\omega}_{m+1} + 2t\hat{\upsilon}_{m+1}). \tag{42}$$

From (34), for $\omega \in (0, 2\pi)$, as $S \to \infty$, we have

$$\frac{1}{S} \sum_{s=0}^{S-1} \exp(j\omega s) = o\left(\frac{\log S}{S}\right)^{\frac{1}{2}}. \tag{43}$$

Since $\frac{|\hat{A}_{m+1}|^2}{2} = \frac{I_u(\hat{\omega}_{m+1}, \hat{\nu}_{m+1})}{ST}$, by (23)

$$|\hat{A}_{m+1}| = O\left(\frac{\log ST}{ST}\right)^{\frac{1}{2}}. \tag{44}$$

Obviously, the amplitudes $\hat{C}_{m+1}$ and $\hat{G}_{m+1}$ have the same order.

From [4], for each $1 \le i \le m$, $(\hat{\omega}_{m+1}, \hat{\nu}_{m+1}) \neq (\hat{\omega}_i, \hat{\nu}_i)$. Therefore, for each $1 \le i \le m$, using (43) and (44), we conclude that

$$M_1 = o\left(\frac{\log ST}{ST}\right). \tag{45}$$

Since $\hat{C}_{m+1}$ and $\hat{G}_{m+1}$ are both of an order $O\left(\frac{\log ST}{ST}\right)^{\frac{1}{2}}$, we have using (43) that the sum of the last two terms of M_2 is of order $o\left(\frac{\log ST}{ST}\right)^{\frac{3}{2}}$. Therefore,

$$\begin{aligned} M_2 &= \frac{|\hat{A}_{m+1}|^2}{2} + o\left(\frac{\log ST}{ST}\right)^{\frac{3}{2}} \\ &= \frac{I_u(\hat{\omega}_{m+1}, \hat{\upsilon}_{m+1})}{ST} + o\left(\frac{\log ST}{ST}\right)^{\frac{3}{2}}. \end{aligned} \tag{46}$$

Substituting (45) and (46) into (40), we have

$$\hat{\sigma}_{m+1}^2 = \hat{\sigma}_m^2 - \frac{I_u(\hat{\omega}_{m+1}, \hat{\upsilon}_{m+1})}{ST} + o\left(\frac{\log ST}{ST}\right). \tag{47}$$

Similarly, one can derive (24).

7. REFERENCES

[1] E. J. Hannan, “Estimating the Dimension of a Linear System,” *J. Multivariate Anal.*, vol. 11, pp. 459-472, 1981.

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

[3] M. Kliger and J. M. Francos, “MAP Model Order Selection Rule for 2-D Sinusoids in White Noise,” accepted for publication, *IEEE Trans. Signal Processing*.

[4] M. Kliger and J. M. Francos, “Strong Consistency of the Over- and Under-Determined LSE of 2-D Exponentials in White Noise,” submitted to *IEEE Trans. Inform. Theory*.

[5] J. Li and P. Stoica, “Efficient Mixed-Spectrum Estimation with Application to Target Feature Extraction,” *IEEE Trans. Signal Processing*, vol. 44, pp. 281-295, 1996.

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

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

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

[9] P. Stoica, P. Eykhoff, P. Janssen and T. Soderstrom, “Model-Structure Selection by Cross-Validation,” *Int. J. Control* , vol. 43, pp. 1841-1878, 1986.