IIT Kanpur

MS Project(2016-17 1^{st} sem)

Estimating model order for 1-D and 2-D cisoid models using adaptively penalizing likelihood rule: Large sample consistency properties

Anupreet Porwal
(12817143)

Supervisors:
Prof. Amit Mitra
Prof. Sharmishtha Mitra

November 3, 2016

Certificate

It is certified that the work embodied in this project entitled "Estimating model order for 1-D and 2-D cisoid models using adaptively penalizing likelihood rule: Large sample consistency properties" by Anupreet Porwal has been carried out under our supervision.

(Amit Mitra)
Professor
Department of Mathematics and Statistics
Indian Institute of Technology Kanpur

(Sharmishtha Mitra)
Associate Professor
Department of Mathematics and Statistics
Indian Institute of Technology Kanpur

Declaration

I hereby declare that the work presented in this project entitled “Estimating model order for 1-D and 2-D cisoid models using adaptively penalizing likelihood rule: Large sample consistency properties” contains my own ideas in my own words. At places, where ideas and words are borrowed from other sources, proper references, as applicable, have been cited. To the best of my knowledge this work does not emanate or resemble to other work created by person(s) other than mentioned herein.

This work was created on November 3, 2016.

-Anupreet Porwal

Acknowledgement

I want to express my sincere gratitude towards my supervisors Prof. Amit Mitra and Prof. Sharmishtha Mitra whose aspiring guidance, unfailing support and continuous efforts and encouragement helped me to learn through out the semester and overcome the difficulties faced throughout the project. I am gratefully indebted to them for helping me acquire a decent knowledge in the concerned topic.

-Anupreet Porwal
November 3, 2016

Abstract

In this project, we look at a novel method for model order selection based on penalizing adaptively the likelihood (PAL) which was recently introduced in [11].Recently, A.Mitra, S.Mita, P.Stoica and K.Surana established the consistency of estimator of model order using PAL rule for nonlinear sinusoidal models in [12]. Here we investigate the asymptotic statistical properties of the same for 1-D and 2-D cisoid models under the assumption of circularly symmetric gaussian error distribution and prove that the estimator is consistent. We present simulation examples to compare the performance of PAL rule with three other information criterion rules for a 1-D cisoid model and look at its behaviour with varying sample size and variance of error distribution.

1 Introduction

Model order selection is a fundamental task in time series analysis and signal processing as once these integer valued parameters are determined, we now know the complexity of the model and the other real valued parameters used to define the model can be estimated. Some of the order selection rules that are associated with the maximum likelihood method of parameter estimation are discussed in [8].Some of the popular rules used for estimation of model order have the following forms:

$$-2\ln f_m(y,\hat{\theta}_m)+\rho(n)m \tag{1}$$

where $\rho(n,m)$ denotes the penalty associated with model order which may depend on sample size n and m where m denotes the model order and $f_m(y,\hat{\theta}_m)$ denotes the probability density function under the hypothesis that y is generated from a model with maximum likelihood estimate of its parameter vector equal to $\hat{\theta}_m \in \mathbb{R}^{m\text{x}1}$. The different penalty terms gives rise to different order selection rules:

$$AIC:\rho(n,m)=2 \tag{2}$$

$$BIC:\rho(n,m)=\ln n \tag{3}$$

$$GIC:\rho(n,m)=1+\rho \text{ where } \rho\geq 1 \tag{4}$$

In this project, we discuss a novel method introduced in [11] based on using a data adaptive penalty and having "oracle-like" properties for 1-D and 2-D cisoid models in section 2 and 3 respectively and prove that the estimators of model order using PAL rules is consistent. In section 4 , we look at a numerical example for a 1-D cisoid model and compare PAL's performance with other widely used information criterion rules followed by conclusion in section 6.

2 PAL rule consistency for 1D Complex cisoid model

We consider the problem of estimating the number of components of the following complex cisoid model
$\forall t=1,2,...,n$

$$y_t=f(t,\theta_m)+\epsilon_t \tag{5}$$

$$f(t,\theta_m)=\sum_{k=1}^{m}\alpha_k e^{it\omega_k} \tag{6}$$

Therefore, the model becomes

$$y_t=\sum_{k=1}^{m}\alpha_k e^{it\omega_k}+\epsilon_t \tag{7}$$

where $\theta_m=(\alpha_{1_R},\alpha_{1_C},\omega_1,...,\alpha_{m_R},\alpha_{m_C},\omega_m)'$ is a $3m$x1 vector of unknown signal parameters; α_{j_R} and α_{j_C} denotes the real and the imaginary part of α_j for j=1,..., m.
Let m_o be the true number components in the the observed signal. Given a sample of size n, $y=(y_1,y_2,...,y_n)'$, the model order selection problem will be to estimate m_o.
Assumptions

- ϵ_t are i.i.d complex valued gaussian with zero mean s.t.

$$\epsilon_t=\epsilon_{t_R}+i\epsilon_{t_C} \tag{8}$$

$$\epsilon_{t_R}\sim\mathcal{N}(0,\sigma^2/2) \tag{9}$$

$$\epsilon_{t_C}\sim\mathcal{N}(0,\sigma^2/2) \tag{10}$$

  and ϵ_{t_R} and ϵ_{t_C} are independent.

- $\forall k = 1, 2, ..., m_o : \omega_k \in (0, 2\pi)$; $\omega_j \neq \omega_k, \forall j \neq k$. Furthermore, $\forall k = 1, 2..., m_o : \alpha_k$ are bounded.
- The true model parameter vector θ_{m_o} is an interior point in the parameter space $\Theta \subset \mathbb{R}^{3m_o}$.

For the model in (7), define

$$\begin{aligned} y &= (y_1, y_2, ..., y_n)' \\ \epsilon &= (\epsilon_1, \epsilon_2, ..., \epsilon_n)' \\ f(\theta_m) &= (f(1, \theta_m), f(2, \theta_m), ..., f(n, \theta_m))'; \end{aligned}$$

For a m- component model, let $\theta_m^* = (\theta_m', \sigma_m^2)'$ denotes the vector containing the underlying signal and noise parameters, then the likelihood function of y under these assumptions can be written as

$$f_m(y, \theta_m^*) = \frac{1}{\left(\frac{2\pi\sigma^2}{2}\right)^n} e^{\frac{-(y-f(\theta_m))^H(y-f(\theta_m))}{\sigma_m^2}} \tag{11}$$

We consider the set of $\tilde{m}$ nested models given by $\{M_m\}_{m=1}^{\tilde{m}}$ where M_m is the m component cisoid model with parameter vector θ_m^*. We assume that the true model M_{m_o} is contained in this set .i.e., $m_o \leq \tilde{m}$ and y is not completely a white noise process i.e., $M_0 \neq M_{m_o}$.
Consider the two Generalized likelihood ratios,

$$r_m = 2\ln\left[\frac{f_{m-1}(y, \hat{\theta}_{m-1}^*)}{f_0(y, \hat{\theta}_0^*)}\right] \tag{12}$$

$$\rho_m = 2\ln\left[\frac{f_{\tilde{m}}(y, \hat{\theta}_{\tilde{m}}^*)}{f_{m-1}(y, \hat{\theta}_{m-1}^*)}\right] \tag{13}$$

where $\hat{\theta}_k^*$ is maximum likelihood estimate of underlying signal and noise parameter vector θ_k^* and $f_0(y, \hat{\theta}_0^*)$ denotes the p.d.f of y when M_0 is the model i.e., $f(\theta_0) = 0$.
The PAL rule can then be defined using the GLR ratios as follows:

$$PAL(m) = -2\ln(f_m(y, \hat{\theta}_m^*)) + (3m+1)\ln(3\tilde{m}+1)\frac{\ln(r_m+1)}{\ln(\rho_m+1)} \tag{14}$$

$$\hat{m} = \underset{m \in \{1,2,...,m\}}{\arg\min} \; [PAL(m)] \tag{15}$$

Remark 1: The ratio $\frac{\ln(r_m+1)}{\ln(\rho_m+1)}$ satisfies:

- increasing function of m
- at m=1, r_1=0 hence the ratio is 0.
- ratio $> 0 \; \forall m \geq 2$

Let us write the model in (7) in matrix form using the following definitions:
$\forall j = 1, 2, ..., m$

$$e_j = (e^{i\omega_j}, e^{i2\omega_j}, ..., e^{in\omega_j})' \tag{16}$$
$$\omega = (\omega_1, \omega_2, ..., \omega_m)' \tag{17}$$
$$D_m(\omega) = (e_1, e_2, ..., e_m)_{\text{nxm}} \tag{18}$$
$$a_m = (\alpha_1, \alpha_2, ..., \alpha_m)'_{\text{mx1}} \tag{19}$$

Under the above notation model in (7) can be written as:

$$y = D_m(\omega)a_m + \epsilon \tag{20}$$

For notational simplicity, we write D_m for $D_m(\omega)$ and $\hat{D_m}$ for $D_m(\hat{\omega})$.Note that in the next sections, $O(.)$ and $o(.)$ denote either the order in probability or deterministic order depending on the context To prove the consistency of the PAL rule

for complex cisoid model we need the proof of following lemmas:
Lemma 1: Under the assumptions A1-A3, $\forall m \leq m_o$,

$$\hat{\sigma}_m^2 = \sigma^2 + \sum_{j=1}^{m_o} \alpha_j^H \alpha_j - \sum_{j=1}^{m} \hat{\alpha}_j^H \hat{\alpha}_j + o(1) \tag{21}$$

a.s. as $n \to \infty$.
Proof: The estimated variance of noise from the model in (20) can be written as:

$$\hat{\sigma}_m^2 = \frac{y^H(I_n - P_m(\hat{\omega}))y}{n} \tag{22}$$

$$\hat{\sigma}_m^2 = \frac{y^H y}{n} - \frac{y^H P_m(\hat{\omega}) y}{n} \tag{23}$$

where $P_m(\hat{\omega}) = \hat{D_m}(\hat{D_m}^H \hat{D_m})^{-1}\hat{D_m}^H$ is the projection matrix. Substituting value of y from true model $y = D_{m_o}a_{m_o} + \epsilon$ and expanding gives:

$$\hat{\sigma}_m^2 = \frac{(D_{m_o}a_{m_o} + \epsilon)^H(D_{m_o}a_{m_o} + \epsilon) - (D_{m_o}a_{m_o} + \epsilon)^H P_m(\hat{\omega})(D_{m_o}a_{m_o} + \epsilon)}{n} \tag{24}$$

$$\hat{\sigma}_m^2 = \frac{1}{n}\left[\epsilon^H\epsilon + 2\epsilon^H D_{m_o}a_{m_o} + a_{m_o}^H D_{m_o}^H D_{m_o}a_{m_o} - \hat{a}_{m_o}^H \hat{D}_{m_o}^H \hat{D}_{m_o}\hat{a}_{m_o}\right] \tag{25}$$

where $\hat{a}_m = (\hat{D_m}^H \hat{D_m})^{-1}\hat{D_m}^H y$. Let us analyze each of the terms one by one

$$\frac{\epsilon^H\epsilon}{n} = \frac{1}{n}\sum_{t=1}^{n}(\epsilon_{t_R}^2 + \epsilon_{t_C}^2) \tag{26}$$

$$= \frac{\sum_{t=1}^{n}\epsilon_{t_R}^2}{n} + \frac{\sum_{t=1}^{n}\epsilon_{t_C}^2}{n} \tag{27}$$

$$\tag{28}$$

Since both ϵ_{t_R} and ϵ_{t_C} are independent, by Kolmogorov's Strong law of Large Numbers(see e.g. [1])

$$\frac{\epsilon^H\epsilon}{n} = \frac{\sum_{t=1}^{n}\epsilon_{t_R}^2}{n} + \frac{\sum_{t=1}^{n}\epsilon_{t_C}^2}{n} \to \frac{\sigma^2}{2} + \frac{\sigma^2}{2} = \sigma^2 \tag{29}$$

a.s. as $n \to \infty$.
Now let us have a look at

$$\frac{\epsilon^H D_{m_o}a_{m_o}}{n} = \frac{1}{n}\begin{bmatrix}\bar{\epsilon}_1 & \bar{\epsilon}_2 & \dots & \bar{\epsilon}_n\end{bmatrix}\begin{bmatrix} e^{i\omega_1} & e^{i\omega_2} & \dots & e^{i\omega_{m_o}} \\ e^{i2\omega_1} & e^{i2\omega_2} & \dots & e^{i2\omega_{m_o}} \\ \vdots & \vdots & \ddots & \vdots \\ e^{in\omega_1} & e^{in\omega_2} & \dots & e^{in\omega_{m_o}}\end{bmatrix}\begin{bmatrix}\alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_{m_o}\end{bmatrix} \tag{30}$$

$$= \frac{1}{n}\begin{bmatrix}\bar{\epsilon}_1 & \bar{\epsilon}_2 & \dots & \bar{\epsilon}_n\end{bmatrix}\begin{bmatrix}\sum_{j=1}^{m_o}\alpha_j e^{i\omega_j} \\ \sum_{j=1}^{m_o}\alpha_j e^{i2\omega_j} \\ \vdots \\ \sum_{j=1}^{m_o}\alpha_j e^{in\omega_j}\end{bmatrix} \tag{31}$$

$$= \frac{1}{n}\sum_{k=1}^{n}\left(\bar{\epsilon}_k\left(\sum_{j=1}^{m_o}\alpha_j e^{ik\omega_j}\right)\right) \tag{32}$$

$$= \frac{1}{n}\left(\sum_{j=1}^{m_o}\alpha_j\left(\sum_{k=1}^{n}\bar{\epsilon}_k e^{ik\omega_j}\right)\right) \tag{33}$$

$$= \frac{1}{n}\left(\sum_{j=1}^{m_o}\alpha_j\left(\sum_{k=1}^{n}\epsilon_{k_R}cos(k\omega_j) + \epsilon_{k_C}sin(k\omega_j) + i\Big(\epsilon_{k_R}sin(k\omega_j) - \epsilon_{k_C}cos(k\omega_j)\Big)\right)\right) \tag{34}$$

Using lemma 1 of Kundu and Mitra [2, 3],

$$\frac{\epsilon^H D_{m_o} a_{m_o}}{n} = o(1) \text{ a.s.} \tag{35}$$

Looking into the third term of (25),

$$\frac{1}{n} a_{m_o}^H D_{m_o}^H D_{m_o} a_{m_o} = \frac{1}{n} \begin{bmatrix} \alpha_1^H & \alpha_2^H & \dots \alpha_{m_o}^H \end{bmatrix} \begin{bmatrix} e^{-i\omega_1} & e^{-i2\omega_1} & \dots & e^{-in\omega_1} \\ e^{-i\omega_2} & e^{-i2\omega_2} & \dots & e^{-in\omega_2} \\ \vdots & \vdots & \ddots & \vdots \\ e^{-i\omega_{m_o}} & e^{-i2\omega_{m_o}} & \dots & e^{-in\omega_{m_o}} \end{bmatrix} \begin{bmatrix} e^{i\omega_1} & e^{i\omega_2} & \dots & e^{i\omega_{m_o}} \\ e^{i2\omega_1} & e^{i2\omega_2} & \dots & e^{i2\omega_{m_o}} \\ \vdots & \vdots & \ddots & \vdots \\ e^{in\omega_1} & e^{in\omega_2} & \dots & e^{in\omega_{m_o}} \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_{m_o} \end{bmatrix} \tag{36}$$

$$= \begin{bmatrix} \alpha_1^H & \alpha_2^H & \dots \alpha_{m_o}^H \end{bmatrix} \begin{bmatrix} 1 & o(1) & \dots & o(1) \\ o(1) & 1 & \dots & o(1) \\ \vdots & \vdots & \ddots & \vdots \\ o(1) & o(1) & \dots & 1 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_{m_o} \end{bmatrix} \tag{37}$$

$$= \sum_{j=1}^{m_o} \alpha_j^H \alpha_j + o(1) \tag{38}$$

$$\frac{1}{n} a_{m_o}^H D_{m_o}^H D_{m_o} a_{m_o} = \sum_{j=1}^{m_o} \alpha_j^H \alpha_j + o(1) \text{ a.s. as } n \to \infty \tag{39}$$

As for $k \neq j$,

$$\lim_{n\to\infty} \frac{\sum_{t=1}^n e^{it(\omega_j - \omega_k)}}{n} = \lim_{n\to\infty} \frac{\sum_{t=1}^n e^{it\tilde{\omega}}}{n} = o(1) \text{ for some fixed } \tilde{\omega} \in (0, 2\pi) \tag{40}$$

Similarly, $\hat{\theta}_m$ is M.L.E. of θ_m, we have

$$\hat{a}_{m_o}^H \hat{D}_{m_o}^H \hat{D}_{m_o} \hat{a}_{m_o} = \sum_{j=1}^{m} \hat{\alpha}_j^H \hat{\alpha}_j + o(1) \text{ a.s. as } n \to \infty \tag{41}$$

Thus using (29), (35), (39) and (41), we hae $\forall m \leq m_o$

$$\hat{\sigma}_m^2 = \sigma^2 + \sum_{j=1}^{m_o} \alpha_j^H \alpha_j - \sum_{j=1}^{m} \hat{\alpha}_j^H \hat{\alpha}_j + o(1) \text{ a.s. as } n \to \infty \tag{42}$$

Lemma 2: Under A1-A3, for any integer $k \geq 1$

$$\hat{\sigma}_{m_o+k}^2 = \hat{\sigma}_{m_o}^2 - \sum_{j=1}^{k} \frac{I_\epsilon(\hat{\omega}_{m_o+j})}{n} + o\left(\frac{\ln n}{n}\right) \text{ a.s. as } n \to \infty \tag{43}$$

or

$$\hat{\sigma}_{m_o+k}^2 = \hat{\sigma}_{m_o}^2 - \frac{G_k}{n} + o\left(\frac{\ln n}{n}\right) \text{ a.s. as } n \to \infty \tag{44}$$

where $G_k = \sum_{j=1}^k I_\epsilon(\hat{\omega}_{m_o+j})$
and

$$I_\epsilon(\omega) = \frac{1}{n} \left| \sum_{t=1}^n \epsilon_t e^{-it\omega} \right|^2 \tag{45}$$

$I_\epsilon(\omega)$ corresponds to periodogram of underlying white noise process and $\hat{\omega}_{m_o+1}, \hat{\omega}_{m_o+2}, ..., \hat{\omega}_{m_o+k}$ are the k largest frequencies corresponding to $I_\epsilon(\omega)$. Thus, G_k is the sum of k largest elements of the periodogram of noise.
Proof: For l=1,2,..., k let

$$\hat{\alpha}_{m_o+l} = \frac{1}{n} \sum_{t=1}^n \epsilon_t e^{-it\hat{\omega}_{m_o+l}} \tag{46}$$

$$\hat{D}_{m_o+k} = (\hat{e}_1, \hat{e}_2, \dots, \hat{e}_{m_o+k}) \tag{47}$$

$$= (\hat{D}_{m_o}, \hat{e}_{m_o+1}, \hat{e}_{m_o+2}, \dots, \hat{e}_{m_o+k}) \tag{48}$$

Similarly, (49)

$$\hat{a}_{m_o+k} = (\hat{a}_{m_o}', \hat{\alpha}_{m_o+1}, \hat{\alpha}_{m_o+2}, \dots, \hat{\alpha}_{m_o+k})' \tag{50}$$

Using result presented in theorem 2 of [4] (see also [5], [6]), we note that in the overestimated scenario, M.L.E of the parameter vector contains a sub vector equal in dimension to the true model order, that converges almost surely to the true parameter vector. Also, the frequencies of the over estimated components are those that maximise the noise periodogram. Thus for the variance of the over estimated model can be given as,

$$\hat{\sigma}^2_{m_o+k} = \frac{y^H(I_n - P_{m_o+k}(\hat{\omega}_{m_o+k}))y}{n} \tag{51}$$

$$\hat{\sigma}^2_{m_o+k} = \frac{y^H y}{n} - \frac{y^H P_{m_o+k}(\hat{\omega}_{m_o+k})y}{n} \tag{52}$$

Substituting the values of $\hat{a}_{m_o+k}$ and $\hat{D}_{m_o+k}$ in the above expressions and simplifying we get

$$\hat{\sigma}^2_{m_o+k} = \frac{1}{n}(y^H y - \hat{a}^H_{m_o}\hat{D}^H_{m_o}\hat{D}_{m_o}\hat{a}_{m_o}) - \frac{2}{n}\sum_{j=1}^{k}\delta_j\Delta_j^H\hat{D}_{m_o}\hat{a}_{m_o} - \frac{1}{n}\left(\sum_{j=1}^{k}\delta_j\Delta_j^H\Delta_j\delta_j^H + 2\sum_{j=1}^{k}\sum_{\substack{\tilde{j}=1\\ j\neq\tilde{j}}}^{k}\delta_j\Delta_j^H\Delta_{\tilde{j}}\delta_{\tilde{j}}^H\right) \tag{53}$$

$$\hat{\sigma}^2_{m_o+k} = \hat{\sigma}^2_{m_o} - T_1 - T_2 \tag{54}$$

where,

$$\hat{\sigma}^2_{m_o} = \frac{1}{n}(y^H y - \hat{a}^H_{m_o}\hat{D}^H_{m_o}\hat{D}_{m_o}\hat{a}_{m_o}) \tag{55}$$

$$T_1 = \frac{2}{n}\sum_{j=1}^{k}\delta_j\Delta_j^H\hat{D}_{m_o}\hat{a}_{m_o} = 2\sum_{k=1}^{k}T_{1_j} \tag{56}$$

$$T_2 = \frac{1}{n}\left(\sum_{j=1}^{k}\delta_j\Delta_j^H\Delta_j\delta_j^H + 2\sum_{j=1}^{k}\sum_{\substack{\tilde{j}=1\\ j\neq\tilde{j}}}^{k}\delta_j\Delta_j^H\Delta_{\tilde{j}}\delta_{\tilde{j}}^H\right) \tag{57}$$

$$= \sum_{j=1}^{k}T_{2_j} + 2\sum_{j=1}^{k}\sum_{\substack{\tilde{j}=1\\ j\neq\tilde{j}}}^{k}T_{2(j,\tilde{j})} \tag{58}$$

$$T_{1_j} = \frac{1}{n}\delta_j\Delta_j^H\hat{D}_{m_o}\hat{a}_{m_o} \tag{59}$$

$$T_{2_j} = \frac{1}{n}\delta_j\Delta_j^H\Delta_j\delta_j^H \tag{60}$$

$$T_{2(j,\tilde{j})} = \frac{1}{n}\delta_j\Delta_j^H\Delta_{\tilde{j}}\delta_{\tilde{j}}^H \tag{61}$$

$$\text{and } \delta_j = \hat{\alpha}_{m_o+j} \text{ and } \Delta_j = \hat{e}_{m_o+j} \tag{62}$$

If ϵ_t is circularly symmetric gaussian with zero mean and finite variance then, $E(\epsilon_{1_R}^2 log|\epsilon_{1_R}|) < \infty$ and $E(\epsilon_{1_C}^2 log|\epsilon_{1_C}|) < \infty$, it thus follows from Theorem 2.2 of [7]:

$$\limsup_{n\to\infty}\frac{\sup_\omega I_\epsilon(\omega)}{\sigma^2\ln(n)} \leq 8 \text{ a.s.} \tag{63}$$

Also, note that $\forall\omega \in (0, 2\pi)$

$$\frac{1}{(n\ln n)^{1/2}}\sum_{t=1}^{n}e^{it\omega} = o(1) \tag{64}$$

which in turn gives us,

$$\frac{1}{n}\sum_{t=1}^{n}e^{it\omega} = o\left(\frac{\ln n}{n}\right)^{1/2} \tag{65}$$

Observe that $\forall j = 1, 2, \ldots, k$

$$|\hat{\alpha}_{m_o+j}|^2 = \frac{1}{n}I_\epsilon(\omega_{\hat{m_o}+j}) \tag{66}$$

$$|\hat{\alpha}_{m_o+j}| = o\left(\frac{\ln n}{n}\right)^{1/2} \text{ a.s.} \tag{67}$$

Using (63),(65) and (66), $\forall j = 1, 2, \ldots, m_o$ as $n \to \infty$

$$\frac{1}{n}\hat{\alpha}_{m_o+1}\hat{e}^H_{m_o+1}\hat{e}_j\hat{\alpha}^H_j = \hat{\alpha}_{m_o+1}\hat{\alpha}^H_j\left(\frac{1}{n}\sum_{t=1}^{n} e^{it(\hat{\omega}_j - \hat{\omega_{m_o+1}})}\right) \tag{68}$$

$$= o\left(\frac{\ln n}{n}\right)^{1/2} o\left(\frac{\ln n}{n}\right)^{1/2} = o\left(\frac{\ln n}{n}\right) \tag{69}$$

Thus,

$$T_{1_1} = \frac{1}{n}\delta_1\Delta_1^H\hat{D}_{m_o}\hat{a}_{m_o} = o\left(\frac{\ln n}{n}\right) \text{ a.s.} \tag{70}$$

or, $\forall j = 1, 2, \ldots, k$

$$T_{1_j} = o\left(\frac{\ln n}{n}\right) \text{ a.s.} \tag{71}$$

$$T_1 = 2\sum_{k=1}^{k} T_{1_j} = o\left(\frac{\ln n}{n}\right)\text{a.s. as } n \to \infty \tag{72}$$

Similarly, $\forall j \neq \tilde{j}$

$$T_{2(j,\tilde{j})} = \delta_j\Delta_j^H\Delta_{\tilde{j}}\delta_{\tilde{j}}^H \tag{73}$$

$$= \frac{1}{n}\hat{\alpha}_{m_o+j}\hat{e}^H_{m_o+j}\hat{e}_{m_o+\tilde{j}}\hat{\alpha}^H_{m_o+\tilde{j}} \tag{74}$$

$$= \hat{\alpha}_{m_o+j}\hat{\alpha}^H_{m_o+\tilde{j}}\left(\frac{1}{n}\hat{e}^H_{m_o+j}\hat{e}_{m_o+\tilde{j}}\right) \tag{75}$$

$$= o\left(\frac{\ln n}{n}\right)^{(3/2)} \tag{76}$$

$$\text{which also implies } T_{2(j,\tilde{j})} = o\left(\frac{\ln n}{n}\right)\text{a.s. as } n \to \infty \tag{77}$$

Also, $\forall 1 \leq j \leq k$,

$$T_{2_j} = \frac{1}{n}\hat{\alpha}_{m_o+j}\hat{e}^H_{m_o+j}\hat{e}_{m_o+j}\hat{\alpha}^H_{m_o+j} \tag{78}$$

$$= \hat{\alpha}_{m_o+j}\hat{\alpha}^H_{m_o+j}\left(\frac{1}{n}\hat{e}^H_{m_o+j}\hat{e}_{m_o+j}\right) \tag{79}$$

$$= |\hat{\alpha}_{m_o+j}|^2 = \frac{1}{n}I_\epsilon(\hat{\omega}_{m_o+j}) \tag{80}$$

Finally substituting value of T_1 and T_2 in (54) we have as $n \to \infty$:

$$\hat{\sigma}^2_{m_o+k} = \hat{\sigma}^2_{m_o} - \sum_{j=1}^{k}\frac{I_\epsilon(\hat{\omega}_{m_o+j})}{n} + o\left(\frac{\ln n}{n}\right) \text{ a.s.} \tag{81}$$

Lemma 3: Under assumptions A1-A3, r_m as defined before in (12) satisfies

$$r_m = \begin{cases} 0 & m = 1 \\ O(n) & 2 \leq m \leq \tilde{m} \end{cases} \tag{82}$$

Proof:

$$r_m = 2\ln\left[\frac{f_{m-1}(y, \hat{\theta}^*_{m-1})}{f_0(y, \hat{\theta}^*_0)}\right] \text{ where,} \tag{83}$$

$$f_m(y, \theta^*_m) = \frac{1}{\left(\frac{2\pi\sigma^2}{2}\right)^n} e^{\frac{-(y-f(\theta_m))^H(y-f(\theta_m))}{\sigma^2_m}} \tag{84}$$

When $\hat{\theta}_m$ is the maximum likelihood estimator of θ_m then $\hat{\sigma}_m^2 = \frac{1}{n}(y - f(\hat{\theta}_m))^H(y - f(\hat{\theta}_m))$.Using this r_m reduces to

$$r_m = 2n \ln\left(\frac{\hat{\sigma}_0^2}{\hat{\sigma}_{m-1}^2}\right) \tag{85}$$

where $\hat{\sigma}_0^2 = \frac{1}{n}\sum_{t=1}^n y_t^H y_t$ as $n \to \infty$. When $m = 1$, clearly $r_1 = 0$. Under normal assumptions of noise, M.L.E $\hat{\theta}_k$ of θ_k is same as the non- linear least sqaure estimate, we conclude using results from [6,5] as $n \to \infty$,

$$\hat{\alpha}_j \to \alpha_j \text{ for j=1,2,\ldots,m; } m \leq m_o \tag{86}$$

Case 1:$m \leq m_o$**(Underestimation)**
$\forall m \leq m_o$,

$$\hat{\sigma}_m^2 \to \sigma^2 + \sum_{j=m+1}^{m_o} \alpha_j^H \alpha_j \text{a.s. and } \hat{\sigma}_{m_o}^2 \to \sigma^2 \text{ a.s. as } n \to \infty \tag{87}$$

$$\text{for } 2 \leq m \leq m_o \tag{88}$$

$$\frac{r_m}{n} = 2\ln\left(\frac{\sigma^2 + \sum_{j=1}^{m_o} \alpha_j^H \alpha_j}{\sigma^2 + \sum_{j=m}^{m_o} \alpha_j^H \alpha_j}\right) \text{ a.s. as } n \to \infty \tag{89}$$

$$= 2\ln\left(1 + \frac{\sum_{j=1}^{m-1} \alpha_j^H \alpha_j}{\sigma^2 + \sum_{j=m}^{m_o} \alpha_j^H \alpha_j}\right) \text{ a.s. as } n \to \infty \tag{90}$$

since the R.H.S is bounded by definition of model and is strictly positive. Hence, we have $r_m = O(n)$ a.s. for all $m \leq m_o$.
Case 2: $m > m_o$**(Overestimation)**
For $m > m_o$, from Lemma 1 and lemma 2 we have

$$\frac{r_m}{n} = 2\ln\left(\frac{\sigma^2 + \sum_{j=1}^{m_o} \alpha_j^H \alpha_j}{\sigma^2 + \frac{G_{m-m_o}}{n} + o\left(\frac{\ln n}{n}\right)}\right) \text{ a.s. as } n \to \infty \tag{91}$$

$$\tag{92}$$

where $G_k = \sum_{j=1}^k I_\epsilon(\hat{\omega}_{m_o+j})$ $\forall k = 1, 2, \ldots, \tilde{m} - m_o$. Using (63), $G_k = O(\ln n)$

$$\frac{G_{m-m_o}}{n} \to 0 \text{ a.s and } o\left(\frac{\ln n}{n}\right) \to 0 \text{ as } n \to \infty \tag{93}$$

Combining this we have for all $m > m_o$

$$\frac{r_m}{n} = 2\ln\left(1 + \frac{\sum_{j=1}^{m_o} \alpha_j^H \alpha_j}{\sigma^2}\right) = O(1) \text{ a.s. as } n \to \infty \tag{94}$$

Hence, we get the desired result.
Lemma4: Under the assumptions A1-A3, ρ_m defined in (13) satisfies

$$\rho_m = \begin{cases} O(n) & m \leq m_o \\ O_p(1) & m = m_o + 1, m_o + 2, \ldots, \tilde{m} \end{cases} \tag{95}$$

Proof:

$$\rho_m = 2\ln\left[\frac{f_{\tilde{m}}(y, \hat{\theta}_{\tilde{m}}^*)}{f_{m-1}(y, \hat{\theta}_{m-1}^*)}\right] \tag{96}$$

$$= 2n\ln\left(\frac{\hat{\sigma}_{m-1}^2}{\hat{\sigma}_{\tilde{m}}^2}\right) \tag{97}$$

Using lemma 2

$$\hat{\sigma}_{\tilde{m}}^2 = \hat{\sigma}_{m_o}^2 - \sum_{j=1}^{\tilde{m}-m_o} \frac{I_\epsilon(\hat{\omega}_{m_o+j})}{n} + o\left(\frac{\ln n}{n}\right) \text{ a.s. as } n \to \infty \tag{98}$$

Also, as $I_\epsilon(\omega) \geq 0 \forall \omega$

$$\frac{1}{n}\sum_{j=1}^{\tilde{m}-m_o} I_\epsilon(\omega_{\hat{m_o+j}}) \leq \frac{\tilde{m}-m_o}{n}\sup_\omega I_\epsilon(\omega) \leq (\tilde{m}-m_o)\frac{\ln n}{n}\frac{\sup_\omega I_\epsilon(\omega)}{\ln n} \to 0 \text{ a.s. as } n \to \infty \tag{99}$$

Case 1:$m \leq m_o$**(Underestimation)**
$\forall m \leq m_o$ as $n \to \infty$

$$\frac{\rho_m}{n} \to 2\ln\left(\frac{\sigma^2 + \sum_{j=m}^{m_o} \alpha_j^H \alpha_j}{\sigma^2}\right) \text{ a.s} \tag{100}$$

The R.H.S is strictly positive and bounded and thus we have

$$\forall m \leq m_o \ \rho_m = O(n) \text{a.s. as} n \to \infty \tag{101}$$

Case 2:$m > m_o$**(overestimation)**
From asymptotic theory of likelihood ratios(see [8, 9, 10]) that as $n \to \infty$

$$\rho_m = 2n\ln\left(\frac{\hat{\sigma}_{m-1}^2}{\hat{\sigma}_{\tilde{m}^2}}\right) \sim \chi^2_{3(\tilde{m}-m+1)} \tag{102}$$

where χ_k^2 denotes the chi square distribution with k degrees of freedom. Therefore, ρ_m is bounded in probability i.e. $O_p(1)$. Using the same we get the desired result.

Remark 2: Using lemma 3 and lemma 4 the ratio $\frac{\ln(r_m+1)}{\ln(\rho_m+1)}$ therefore satisfies

$$\frac{\ln(r_m+1)}{\ln(\rho_m+1)} = \begin{cases} O(1) & m \leq m_o \\ O(\ln n) & m = m_o+1, m_o+2, \ldots, \tilde{m} \end{cases} \tag{103}$$

Theorem: Under A1-A3, if m_o is true order($m_o \leq \tilde{m}$) and $\hat{m}$ is the estimated model order using PAL rule then

$$P(\hat{m} \neq m_o) \to 0 \text{ as } \ n \to \infty \tag{104}$$

Proof: Under assumption of the model the PAL rule mentioned in (14) can also be written as

$$PAL(m) = 2n\ln(\hat{\sigma}_m^2) + (3m+1)\ln(3\tilde{m}+1)\frac{\ln(r_m+1)}{\ln(\rho_m+1)} + \gamma \tag{105}$$

where γ is a constant. **Case 1: Underestimation(**$m \leq m_o$**)**

$$PAL(m) - PAL(m_o) = 2n\ln(\frac{\hat{\sigma}_m^2}{\hat{\sigma}_{m_o}^2}) + (3m+1)\ln(3\tilde{m}+1)\frac{\ln(r_m+1)}{\ln(\rho_m+1)} - (3m_o+1)\ln(3\tilde{m}+1)\frac{\ln(r_{m_o}+1)}{\ln(\rho_{m_o}+1)} \tag{106}$$

Using lemma 1

- $\hat{\sigma}_{m_o}^2 \to \sigma^2$ a.s. as $n \to \infty$
- $\forall m \leq m_o, \ \hat{\sigma}_m^2 \to \sigma^2 + \sum_{j=m+1}^{m_o} \alpha_j^H \alpha_j$ a.s. as $n \to \infty$

$\forall m < m_o$

$$\ln(\frac{\hat{\sigma}_m^2}{\hat{\sigma}_{m_o}^2}) \to 2\ln\left(1 + \frac{\sum_{j=m+1}^{m_o} \alpha_j^H \alpha_j}{\sigma^2}\right) \text{ a.s as} n \to \infty \tag{107}$$

where R.H.S. is strictly positive and bounded by the definition of the model. Also, using lemma 3 and 4 we know that $r_m = O(n)$ and $\rho_m = O(n)$ a.s.

$$(3m+1)\ln(3\tilde{m}+1)\frac{\ln(r_m+1)}{\ln(\rho_m+1)} - (3m_o+1)\ln(3\tilde{m}+1)\frac{\ln(r_{m_o}+1)}{\ln(\rho_{m_o}+1)} \tag{108}$$

$$= (3m+1)\ln(3\tilde{m}+1)\frac{O(\ln n)}{O(\ln n)} - (3m_o+1)\ln(3\tilde{m}+1)\frac{O(\ln n)}{O(\ln n)} \text{ a.s as } n \to \infty \tag{109}$$

$$PAL(m) - PAL(m_o) = O(n) + (3m+1)\ln(3\tilde{m}+1)\frac{O(\ln n)}{O(\ln n)} - (3m_o+1)\ln(3\tilde{m}+1)\frac{O(\ln n)}{O(\ln n)} \ \text{ a.s.} \tag{110}$$

$$\frac{1}{n}(PAL(m) - PAL(m_o)) \to 2\ln\left(1 + \frac{\sum_{j=m+1}^{m_o} \alpha_j^H \alpha_j}{\sigma^2}\right) \text{ a.s.} \tag{111}$$

as $n \to \infty$ where R.H.S. is a strictly positive and bounded quantity.Thus,
$PAL(m) > PAL(m_o) \quad \forall m < m_o$ with probability 1

$$P(\hat{m} < m_o) = P(PAL(m) < PAL(m_o) \text{ for some } m < m_o) \tag{112}$$

$$= P\left(\bigcup_{m<m_o} PAL(m) < PAL(m_o)\right) \tag{113}$$

$$\leq \sum_{m=1}^{m_o-1} P\left(\frac{1}{n}(PAL(m) - PAL(m_o) < 0)\right) \to 0 \tag{114}$$

as $n \to \infty$.

Case 2: Overestimation($m > m_o$) We observe that

$$PAL(m) - PAL(m_o) = 2n\ln(\frac{\hat{\sigma}_m^2}{\sigma_{m_o}^2}) + (3m+1)\ln(3\tilde{m}+1)\frac{\ln(r_m+1)}{\ln(\rho_m+1)} - (3m_o+1)\ln(3\tilde{m}+1)\frac{\ln(r_{m_o}+1)}{\ln(\rho_{m_o}+1)} \tag{115}$$

We know from Remark 1 that $\forall m < m_o$ and for all n

$$\frac{\ln(r_m+1)}{\ln(\rho_m+1)} \geq \frac{\ln(r_{m_o}+1)}{\ln(\rho_{m_o}+1)} \tag{116}$$

as ratio is an increasing function of m.Therefore we can write (115) as

$$\frac{1}{\ln n}\left(PAL(m) - PAL(m_o)\right) \geq \frac{1}{\ln n}\left(2n\ln(\frac{\hat{\sigma}_m^2}{\hat{\sigma}_{m_o}^2}) + 3(m-m_o)\ln(3\tilde{m}+1)\frac{\ln(r_m+1)}{\ln(\rho_m+1)}\right) \quad \forall n \tag{117}$$

$$P\left(\frac{1}{\ln n}\left(PAL(m) - PAL(m_o)\right) > 0\right) = P\left(\frac{1}{\ln n}\left(2n\ln(\frac{\hat{\sigma}_m^2}{\sigma_{m_o}^2})\right) + \frac{1}{\ln n}\left(3(m-m_o)\ln(3\tilde{m}+1)\frac{\ln(r_m+1)}{\ln(\rho_m+1)}\right) > 0\right) \tag{118}$$

It follows from the asymptotic theory of likelihood ratios (see [10, 9]) that,

$$2n\ln\left(\frac{\hat{\sigma}_m^2}{\hat{\sigma}_{m_o}^2}\right) \sim \chi^2_{3(m-m_o)} \tag{119}$$

Thus, $\frac{1}{\ln(n)}\left(2n\ln\left(\frac{\hat{\sigma}_m^2}{\hat{\sigma}_{m_o}^2}\right)\right) = o_p(1)$.Also, $\frac{1}{\ln n}3(m-m_o)\ln(3m+1) > 0$ and the ratio $\frac{1}{\ln n}\left(\frac{\ln(r_m+1)}{\ln(\rho_m+1)}\right) = o(1) \quad \forall n$

$$P\left(\frac{1}{\ln n}\left(2n\ln(\frac{\hat{\sigma}_m^2}{\sigma_{m_o}^2})\right) + \frac{1}{\ln n}\left(3(m-m_o)\ln(3\tilde{m}+1)\frac{\ln(r_m+1)}{\ln(\rho_m+1)}\right) > 0\right) \to 1 \tag{120}$$

Using (120), we have that

$$P(\hat{m} > m_o) = P(PAL(m) < PAL(m_o) \text{ for some } m > m_o) \tag{121}$$

$$= P\left(\bigcup_{m>m_o} PAL(m) < PAL(m_o)\right) \tag{122}$$

$$\leq \sum_{m=m_o+1}^{\tilde{m}} P\left(\frac{1}{\ln n}(PAL(m) - PAL(m_o) < 0)\right) \to 0 \tag{123}$$

as $n \to \infty$. This implies

$$P(\hat{m} \neq m_o) \to 0 \text{ as } n \to \infty \tag{124}$$

3 PAL rule consistency for 2D Complex cisoid model

We now consider the problem of estimating the model order for the following 2-dimensional complex cisoid model $\forall s, t$ such that$1 \leq s \leq S; 1 \leq t \leq T$

$$y(s,t) = f(s,t,\theta_m) + \epsilon(s,t) \tag{125}$$

$$f(s,t,\theta_m) = \sum_{k=1}^{m} \alpha_k e^{i(s\beta_k + t\omega_k)} \tag{126}$$

Therefore, the model becomes

$$y(s,t) = \sum_{k=1}^{m} \alpha_k e^{i(s\beta_k + t\omega_k)} + \epsilon(s,t) \tag{127}$$

where $\theta_m = (\alpha_{1_R}, \alpha_{1_C}, \beta_1, \omega_1, ..., \alpha_{m_R}, \alpha_{m_C}, \beta_m, \omega_m)'$ is a $4m$x1 vector of unknown signal parameters; α_{j_R} and α_{j_C} denotes the real and the imaginary part of α_j for j=1,..., m.
Let m_o be the true number of components in the the observed signal. Given a sample of size ST,

$$y = (y(1,1), y(2,1), \ldots, y(S,1), y(1,2), \ldots, y(S,2), \ldots, y(S,T))'$$

the model order selection problem will be to estimate m_o.
Assumptions

- $\epsilon(s,t)$ are i.i.d complex valued gaussian with zero mean s.t.

$$\epsilon(s,t) = \epsilon_R(s,t) + i\epsilon_C(s,t) \tag{128}$$

$$\epsilon_R(s,t) \sim \mathcal{N}(0, \sigma^2/2) \tag{129}$$

$$\epsilon_C(s,t) \sim \mathcal{N}(0, \sigma^2/2) \tag{130}$$

  and $\epsilon_R(s,t)$ and $\epsilon_C(s,t)$ are independent.

- $\forall k = 1, 2, ..., m_o : (\beta_k, \omega_k) \in (0, 2\pi)x(0, 2\pi)$; where (β_k, ω_k) are pairwise different i.e. $\omega_j \neq \omega_k$ or $\beta_j \neq \beta_k, \forall j \neq k$. Furthermore, $\forall k = 1, 2..., m_o : \alpha_k$ are bounded.

- The true model parameter vector θ_{m_o} is an interior point in the parameter space $\Theta \subset \mathbb{R}^{4m_o}$.

For the model in (127), define

$$\begin{aligned} y &= (y(1,1), y(2,1), \ldots, y(S,1), y(1,2), \ldots, y(S,2), \ldots, y(S,T))' \\ \epsilon &= (\epsilon(1,1), \epsilon(2,1), \ldots, \epsilon(S,1), \epsilon(1,2), \ldots, \epsilon(S,2), \ldots, \epsilon(S,T))' \\ f(\theta_m) &= (f(1,1,\theta_m), f(2,1,\theta_m), \ldots, f(S,1,\theta_m), f(1,2,\theta_m), \ldots, f(S,2,\theta_m), \ldots, f(S,T,\theta_m))'; \end{aligned}$$

For a m- component model, let $\theta_m^* = (\theta_m', \sigma_m^2)'$ denotes the vector containing the underlying signal and noise parameters, then the likelihood function of y under these assumptions can be written as

$$f_m(y, \theta_m^*) = \frac{1}{\left(\frac{2\pi\sigma^2}{2}\right)^{ST}} e^{\frac{-(y-f(\theta_m))^H (y-f(\theta_m))}{\sigma_m^2}} \tag{131}$$

We consider the set of $\tilde{m}$ nested models given by $\{M_m\}_{m=1}^{\tilde{m}}$ where M_m is the m component 2-D cisoid model with parameter vector θ_m^*. We assume that the true model M_{m_o} is contained in this set .i.e., $m_o \leq \tilde{m}$ and y is not completely a white noise process i.e., $M_0 \neq M_{m_o}$.
Using the two Generalized likelihood ratios defined in (12) and (13), the PAL rule for this model can then be defined as follows:

$$PAL(m) = -2\ln(f_m(y, \hat{\theta}_m^*)) + (4m+1)\ln(4\tilde{m}+1)\frac{\ln(r_m+1)}{\ln(\rho_m+1)} \tag{132}$$

$$\hat{m} = \underset{m \in \{1,2,\ldots,m\}}{\arg\min} [PAL(m)] \tag{133}$$

Let us write the model in (127) in matrix form using the following definitions:
$\forall j = 1, 2, ..., m$

$$e_j = (e^{i(\beta_k+\omega_k)}, e^{i(2\beta_k+\omega_k)}, \ldots, e^{i(S\beta_k+\omega_k)}, e^{i(1\beta_k+2\omega_k)}, \ldots, e^{i(S\beta_k+2\omega_k)}, \ldots, e^{i(S\beta_k+T\omega_k)})' \tag{134}$$
$$\beta = (\beta_1, \beta_2, ..., \beta_m)' \tag{135}$$
$$\omega = (\omega_1, \omega_2, ..., \omega_m)' \tag{136}$$
$$D_m(\beta, \omega) = (e_1, e_2, ..., e_m)_{STxm} \tag{137}$$
$$a_m = (\alpha_1, \alpha_2, ..., \alpha_m)'_{mx1} \tag{138}$$

Under the above notation model in 7 can be written as:

$$y = D_m(\beta, \omega)a_m + \epsilon \tag{139}$$

For notational simplicity, we write D_m for $D_m(\beta, \omega)$ and $\hat{D_m}$ for $D_m(\hat{\beta}, \hat{\omega})$.Note that in the next sections, $O(.)$ and $o(.)$ denote either the order in probability or deterministic order depending on the context.
Remark: Let $\{\Psi_i\}$ be a sequence of rectangles s.t.

$$\Psi = \{(s,t) \in \mathbb{Z}^2 | 1 \le s \le S_i, 1 \le t \le T_i\} \tag{140}$$

Then, sequence of subsets $\{\Psi_i\}$ is said to tend to infinity as $i \to \infty$ if

$$\lim_{i\to\infty} min(S_i, T_i) \to \infty \text{ and } 0 < \lim_{i\to\infty} \left(\frac{S_i}{T_i}\right) < \infty \tag{141}$$

To simplify notation, omit subscript i. Thus, $\Psi(S,T) \to \infty$ implies both S and T tend to infinity as a function of i, and roughly at the same rate.

To prove the consistency of the PAL rule for 2-D complex cisoid model we need the proof of following lemmas:
Lemma 1: Under the assumptions A1-A3, $\forall m \le m_o$,

$$\hat{\sigma}^2_m = \sigma^2 + \sum_{j=1}^{m_o} \alpha_j^H \alpha_j - \sum_{j=1}^{m} \hat{\alpha}_j^H \hat{\alpha}_j + o(1) \tag{142}$$

a.s. as $\Psi(S,T) \to \infty$.
Proof: The estimated variance of noise from the model in 139 can be written as:

$$\hat{\sigma}^2_m = \frac{y^H(I_{ST} - P_m(\hat{\beta}, \hat{\omega}))y}{ST} \tag{143}$$
$$\hat{\sigma}^2_m = \frac{y^H y}{ST} - \frac{y^H P_m(\hat{\beta}, \hat{\omega})y}{ST} \tag{144}$$

where $P_m(\hat{\beta}, \hat{\omega}) = \hat{D_m}(\hat{D_m}^H \hat{D_m})^{-1}\hat{D_m}^H$ is the projection matrix. Substituting value of y from true model $y = D_{m_o}a_{m_o} + \epsilon$ and expanding gives:

$$\hat{\sigma}^2_m = \frac{(D_{m_o}a_{m_o} + \epsilon)^H(D_{m_o}a_{m_o} + \epsilon) - (D_{m_o}a_{m_o} + \epsilon)^H P_m(\hat{\omega})(D_{m_o}a_{m_o} + \epsilon)}{ST} \tag{145}$$

$$\hat{\sigma}^2_m = \frac{1}{ST}\left[\epsilon^H\epsilon + 2\epsilon^H D_{m_o}a_{m_o} + a^H_{m_o}D^H_{m_o}D_{m_o}a_{m_o} - \hat{a}^H_{m_o}\hat{D}^H_{m_o}\hat{D}_{m_o}\hat{a}_{m_o}\right] \tag{146}$$

where $\hat{a}_m = (\hat{D_m}^H \hat{D_m})^{-1}\hat{D_m}^H y$. Let us analyze each of the terms one by one

$$\frac{\epsilon^H\epsilon}{ST} = \frac{1}{ST}\sum_{s=1}^{S}\sum_{t=1}^{T}(\epsilon_R^2(s,t) + \epsilon_C^2(s,t)) \tag{147}$$
$$= \frac{\sum_{s=1}^{S}\sum_{t=1}^{T}\epsilon_R^2(s,t)}{ST} + \frac{\sum_{s=1}^{S}\sum_{t=1}^{T}\epsilon_C^2(s,t)}{ST} \tag{148}$$

(149)

Since both $\epsilon_R(s,t)$ and $\epsilon_C(s,t)$ are independent, by Kolmogorov's Strong law of Large Numbers(see e.g. [1])

$$\frac{\epsilon^H \epsilon}{ST} = \frac{\sum_{s=1}^{S}\sum_{t=1}^{T}\epsilon_R^2(s,t)}{ST} + \frac{\sum_{s=1}^{S}\sum_{t=1}^{T}\epsilon_C^2(s,t)}{ST} \to \frac{\sigma^2}{2} + \frac{\sigma^2}{2} = \sigma^2 \tag{150}$$

a.s. as $\Psi(S,T) \to \infty$.
Now let us have a look at

$$\frac{\epsilon^H D_{m_o} a_{m_o}}{ST} = \frac{1}{ST}\begin{bmatrix}\bar{\epsilon}(1,1) & \bar{\epsilon}(2,1) & \dots & \bar{\epsilon}(S,T)\end{bmatrix}\begin{bmatrix}\sum_{k=1}^{m_o}\alpha_k e^{i(\beta_k+\omega_k)} \\ \sum_{k=1}^{m_o}\alpha_k e^{i(2\beta_k+\omega_k)} \\ \vdots \\ \vdots \\ \sum_{k=1}^{m_o}\alpha_k e^{i(S\beta_k+T\omega_k)}\end{bmatrix} \tag{151}$$

$$= \frac{1}{ST}\sum_{l=1}^{S}\sum_{j=1}^{T}\left(\bar{\epsilon}(l,j)\left(\sum_{k=1}^{m_o}\alpha_k e^{i(l\beta_k+j\omega_k)}\right)\right) \tag{152}$$

$$= \frac{1}{ST}\left(\sum_{k=1}^{m_o}\alpha_k\left(\sum_{l=1}^{S}\sum_{j=1}^{T}\bar{\epsilon}(l,j)e^{i(l\beta_k+j\omega_k)}\right)\right) \tag{153}$$

$$= \frac{1}{ST}\left(\sum_{k=1}^{m_o}\alpha_k\left(\sum_{l=1}^{S}\sum_{j=1}^{T}\epsilon_R(l,j)cos(l\beta_k+j\omega_k) + \epsilon_C(l,j)sin(l\beta_k+j\omega_k) + i\Big(\epsilon_R(l,j)sin(l\beta_k+j\omega_k) - \epsilon_C(l,j)cos(l\beta_k+j\omega_k)\Big)\right)\right) \tag{154}$$

Using lemma 2 of Kundu and Mitra [2, 3],

$$\frac{\epsilon^H D_{m_o} a_{m_o}}{ST} = o(1) \text{ a.s.} \tag{155}$$

Looking into the third term of (146),

$$\frac{1}{ST}a_{m_o}^H D_{m_o}^H D_{m_o} a_{m_o} == \begin{bmatrix}\alpha_1^H & \alpha_2^H & \dots \alpha_{m_o}^H\end{bmatrix}\frac{D_{m_o}^H D_{m_o}}{ST}\begin{bmatrix}\alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_{m_o}\end{bmatrix} \tag{156}$$

$$\tag{157}$$

where (i,j) element of $\frac{D_{m_o}^H D_{m_o}}{ST}$ can be given as,

$$\frac{(D_{m_o}^H D_{m_o})_{ij}}{ST} = \begin{cases} 1 & i = j \\ \left(\frac{1}{S}\sum_{s=1}^{S}e^{is(\beta_j-\beta_i)}\right)\left(\frac{1}{T}\sum_{t=1}^{T}e^{it(\omega_j-\omega_i)}\right) & i \neq j \end{cases} \tag{158}$$

As for $k \neq j$,

$$\lim_{n\to\infty}\frac{\sum_{t=1}^{n}e^{it(\omega_j-\omega_k)}}{n} = \lim_{n\to\infty}\frac{\sum_{t=1}^{n}e^{it\tilde{\omega}}}{n} = o(1) \text{ for some fixed } \tilde{\omega} \in (0,2\pi) \tag{159}$$

Therefore, for $i \neq j$ case both the terms goes to $o(1)$ as $\Psi(S,T) \to \infty$.Thus,

$$\frac{1}{ST}a_{m_o}^H D_{m_o}^H D_{m_o} a_{m_o} = \sum_{j=1}^{m_o}\alpha_j^H \alpha_j + o(1) \text{ a.s. as } \Psi(S,T) \to \infty \tag{160}$$

Similarly, $\hat{\theta}_m$ is M.L.E. of θ_m, we have

$$\frac{1}{ST}\hat{a}_{m_o}^H \hat{D}_{m_o}^H \hat{D}_{m_o}\hat{a}_{m_o} = \sum_{j=1}^{m}\hat{\alpha}_j^H \hat{\alpha}_j + o(1) \text{ a.s. as } \Psi(S,T) \to \infty \tag{161}$$

Thus using (150), (155), (160) and (161), we have $\forall m \leq m_o$

$$\hat{\sigma}^2_m = \sigma^2 + \sum_{j=1}^{m_o} \alpha_j^H \alpha_j - \sum_{j=1}^{m} \hat{\alpha}_j^H \hat{\alpha}_j + o(1) \text{ a.s. as } \Psi(S,T) \to \infty \tag{162}$$

Lemma 2: Under A1-A3, for any integer $k \geq 1$

$$\hat{\sigma}^2_{m_o+k} = \hat{\sigma}^2_{m_o} - \sum_{j=1}^{k} \frac{I_\epsilon(\hat{\beta}_{m_o+j}, \hat{\omega}_{m_o+j})}{ST} + o\left(\frac{(\ln(ST) \ln S \ln T)^{1/2}}{ST} \right) \text{ a.s. as } \Psi(S,T) \to \infty \tag{163}$$

or

$$\hat{\sigma}^2_{m_o+k} = \hat{\sigma}^2_{m_o} - \frac{G_k}{ST} + o\left(\frac{(\ln(ST) \ln S \ln T)^{1/2}}{ST} \right) \text{ a.s. as } \Psi(S,T) \to \infty \tag{164}$$

where $G_k = \sum_{j=1}^{k} I_\epsilon(\hat{\beta}_{m_o+j}, \hat{\omega}_{m_o+j})$
and

$$I_\epsilon(\beta,\omega) = \frac{1}{ST} \left| \sum_{s=1}^{S} \sum_{t=1}^{T} \epsilon(s,t) e^{-i(s\beta + t\omega)} \right|^2 \tag{165}$$

$I_\epsilon(\beta,\omega)$ corresponds to periodogram of underlying white noise process and $(\hat{\beta}_{m_o+1}, \hat{\omega}_{m_o+1}), (\hat{\beta}_{m_o+2}, \hat{\omega}_{m_o+2}), \ldots, (\hat{\beta}_{m_o+k}, \hat{\omega}_{m_o+k})$ are the k largest frequencies corresponding to $I_\epsilon(\beta,\omega)$. Thus, G_k is the sum of k largest elements of the periodogram of noise.
Proof: For l=1,2,..., k let

$$\hat{\alpha}_{m_o+l} = \frac{1}{ST} \sum_{s=1}^{S} \sum_{t=1}^{T} \epsilon(s,t) e^{-i(s\hat{\beta}_{m_o+l} + t\hat{\omega}_{m_o+l})} \tag{166}$$

$$\hat{D}_{m_o+k} = (\hat{e}_1, \hat{e}_2, \ldots, \hat{e}_{m_o+k}) \tag{167}$$

$$= (\hat{D}_{m_o}, \hat{e}_{m_o+1}, \hat{e}_{m_o+2}, \ldots, \hat{e}_{m_o+k}) \tag{168}$$

$$\text{Similarly,} \tag{169}$$

$$\hat{a}_{m_o+k} = (\hat{a}'_{m_o}, \hat{\alpha}_{m_o+1}, \hat{\alpha}_{m_o+2}, \ldots, \hat{\alpha}_{m_o+k})' \tag{170}$$

Using results presented in theorem 2 of [4] (see also [5], [6]) we note that in the overestimated scenario, M.L.E of the parameter vector contains a sub vector equal in dimension to the true model order, that converges almost surely to the true parameter vector. Also, the frequencies of the over estimated components are those that maximise the noise periodogram. Thus for the variance of the over estimated model can be given as,

$$\hat{\sigma}^2_{m_o+k} = \frac{y^H(I_{ST} - P_{m_o+k}(\hat{\beta}_{m_o+k}, \hat{\omega}_{m_o+k}))y}{ST} \tag{171}$$

$$\hat{\sigma}^2_{m_o+k} = \frac{y^H y}{ST} - \frac{y^H P_{m_o+k}(\hat{\beta}_{m_o+k}, \hat{\omega}_{m_o+k}) y}{ST} \tag{172}$$

Substituting the values of $\hat{a}_{m_o+k}$ and $\hat{D}_{m_o+k}$ in the above expressions and simplifying we get

$$\hat{\sigma}^2_{m_o+k} = \frac{1}{ST}(y^H y - \hat{a}^H_{m_o} \hat{D}^H_{m_o} \hat{D}_{m_o} \hat{a}_{m_o}) - \frac{2}{ST} \sum_{j=1}^{k} \delta_j \Delta_j^H \hat{D}_{m_o} \hat{a}_{m_o} - \frac{1}{ST} \left(\sum_{j=1}^{k} \delta_j \Delta_j^H \Delta_j \delta_j^H + 2 \sum_{\substack{j=1 \\ j \neq \tilde{j}}}^{k} \sum_{\tilde{j}=1}^{k} \delta_j \Delta_j^H \Delta_{\tilde{j}} \delta_{\tilde{j}}^H \right) \tag{173}$$

$$\hat{\sigma}^2_{m_o+k} = \hat{\sigma}^2_{m_o} - T_1 - T_2 \tag{174}$$

where,

$$\hat{\sigma}^2_{m_o} = \frac{1}{ST}(y^H y - \hat{a}^H_{m_o}\hat{D}^H_{m_o}\hat{D}_{m_o}\hat{a}_{m_o}) \tag{175}$$

$$T_1 = \frac{2}{ST}\sum_{j=1}^{k}\delta_j\Delta_j^H\hat{D}_{m_o}\hat{a}_{m_o} = 2\sum_{k=1}^{k}T_{1_j} \tag{176}$$

$$T_2 = \frac{1}{ST}\left(\sum_{j=1}^{k}\delta_j\Delta_j^H\Delta_j\delta_j^H + 2\sum_{\substack{j=1\\ j\neq\tilde{j}}}^{k}\sum_{\tilde{j}=1}^{k}\delta_j\Delta_j^H\Delta_{\tilde{j}}\delta_{\tilde{j}}^H\right) \tag{177}$$

$$= \sum_{j=1}^{k}T_{2_j} + 2\sum_{\substack{j=1\\ j\neq\tilde{j}}}^{k}\sum_{\tilde{j}=1}^{k}T_{2(j,\tilde{j})} \tag{178}$$

$$T_{1_j} = \frac{1}{ST}\delta_j\Delta_j^H\hat{D}_{m_o}\hat{a}_{m_o} \tag{179}$$

$$T_{2_j} = \frac{1}{ST}\delta_j\Delta_j^H\Delta_j\delta_j^H \tag{180}$$

$$T_{2(j,\tilde{j})} = \frac{1}{ST}\delta_j\Delta_j^H\Delta_{\tilde{j}}\delta_{\tilde{j}}^H \tag{181}$$

$$\tag{182}$$

and $\delta_j = \hat{\alpha}_{m_o+j}$ and $\Delta_j = \hat{e}_{m_o+j}$

If $\epsilon(s,t)$ is circularly symmetric gaussian with zero mean and finite variance then, $E(\epsilon_R(1,1)^2 log|\epsilon_R(1,1)|) < \infty$ and $E(\epsilon_C(1,1)^2 log|\epsilon_C(1,1)|) < \infty$, it thus follows from Theorem 2.2 of [7]:

$$\limsup_{\Psi(S,T)\to\infty}\frac{\sup_{(\beta,\omega)} I_\epsilon(\beta,\omega)}{\sigma^2\ln(ST)} \leq 8 \text{ a.s.} \tag{183}$$

Also, note that $\forall\omega \in (0,2\pi)$

$$\frac{1}{(n\ln n)^{1/2}}\sum_{t=1}^{n}e^{it\omega} = o(1) \tag{184}$$

which in turn gives us,

$$\frac{1}{n}\sum_{t=1}^{n}e^{it\omega} = o\left(\frac{\ln n}{n}\right)^{1/2} \tag{185}$$

Observe that $\forall j = 1,2,\ldots,k$

$$|\hat{\alpha}_{m_o+j}|^2 = \frac{1}{ST}I_\epsilon(\beta_{\hat{m_o+j}}, \omega_{\hat{m_o+j}}) \tag{186}$$

$$|\hat{\alpha}_{m_o+j}| = o\left(\frac{\ln(ST)}{ST}\right)^{1/2} \text{ a.s.} \tag{187}$$

Using (183),(185) and (186), $\forall j = 1,2,\ldots,m_o$ as $\Psi(S,T)\to\infty$

$$\frac{1}{ST}\hat{\alpha}_{m_o+1}\hat{e}^H_{m_o+1}\hat{e}_j\hat{\alpha}^H_j = \hat{\alpha}_{m_o+1}\alpha^H_j\left(\frac{1}{ST}\sum_{s=1}^{S}\sum_{t=1}^{T}e^{i(s(\hat{\beta}_j-\beta_{\hat{m_o+1}})+t(\hat{\omega}_j-\omega_{\hat{m_o+1}}))}\right) \tag{188}$$

$$= o\left(\frac{\ln(ST)}{ST}\right)^{1/2} o\left(\frac{\ln S}{S}\right)^{1/2} o\left(\frac{\ln T}{T}\right)^{1/2} = o\left(\frac{(\ln(ST)\ln S\ln T)^{1/2}}{ST}\right) \tag{189}$$

Thus,

$$T_{1_1} = \frac{1}{ST}\delta_1\Delta_1^H\hat{D}_{m_o}\hat{a}_{m_o} = o\left(\frac{(\ln(ST)\ln S\ln T)^{1/2}}{ST}\right) \text{ a.s.} \tag{190}$$

or, $\forall j = 1, 2, \ldots, k$

$$T_{1_j} = o\left(\frac{(\ln(ST)\ln S \ln T)^{1/2}}{ST}\right) \text{ a.s.} \tag{191}$$

$$T_1 = 2\sum_{k=1}^{k} T_{1_j} = o\left(\frac{(\ln(ST)\ln S \ln T)^{1/2}}{ST}\right) \text{a.s. as } \Psi(S,T) \to \infty \tag{192}$$

Similarly, $\forall j \neq \tilde{j}$

$$T_{2(j,\tilde{j})} = \frac{1}{ST}\delta_j \Delta_j^H \Delta_{\tilde{j}} \delta_{\tilde{j}}^H \tag{193}$$

$$= \frac{1}{ST}\hat{\alpha}_{m_o+j}\hat{e}_{m_o+j}^H \hat{e}_{m_o+\tilde{j}}\hat{\alpha}_{m_o+\tilde{j}}^H \tag{194}$$

$$= \hat{\alpha}_{m_o+j}\hat{\alpha}_{m_o+\tilde{j}}^H \left(\frac{1}{ST}\hat{e}_{m_o+j}^H \hat{e}_{m_o+\tilde{j}}\right) \tag{195}$$

$$= o\left(\frac{(\ln(ST)\ln S \ln T)^{1/2}}{ST}\right)\text{a.s. as } \Psi(S,T) \to \infty \tag{196}$$

Also, $\forall 1 \leq j \leq k$,

$$T_{2_j} = \frac{1}{ST}\hat{\alpha}_{m_o+j}\hat{e}_{m_o+j}^H \hat{e}_{m_o+j}\hat{\alpha}_{m_o+j}^H \tag{197}$$

$$= \hat{\alpha}_{m_o+j}\hat{\alpha}_{m_o+j}^H \left(\frac{1}{ST}\hat{e}_{m_o+j}^H \hat{e}_{m_o+j}\right) \tag{198}$$

$$= |\hat{\alpha}_{m_o+j}|^2 = \frac{1}{ST} I_\epsilon(\beta_{m_o+j}^{\hat{}}, \omega_{m_o+j}^{\hat{}}) \tag{199}$$

Finally substituting value of T_1 and T_2 in (174) we have as $\Psi(S,T) \to \infty$:

$$\hat{\sigma}_{m_o+k}^2 = \hat{\sigma}_{m_o}^2 - \sum_{j=1}^{k} \frac{I_\epsilon(\hat{\beta}_{m_o+j}, \hat{\omega}_{m_o+j})}{ST} + o\left(\frac{(\ln(ST)\ln S \ln T)^{1/2}}{ST}\right) \text{ a.s.} \tag{200}$$

Lemma 3: Under assumptions A1-A3, r_m as defined before in (12) satisfies

$$r_m = \begin{cases} 0 & m = 1 \\ O(ST) & 2 \leq m \leq \tilde{m} \end{cases} \tag{201}$$

Proof:

$$r_m = 2\ln\left[\frac{f_{m-1}(y, \hat{\theta}_{m-1}^*)}{f_0(y, \hat{\theta}_0^*)}\right] \text{ where,} \tag{202}$$

$$f_m(y, \theta_m^*) = \frac{1}{\left(\frac{2\pi\sigma^2}{2}\right)^{ST}} e^{\frac{-(y-f(\theta_m))^H (y-f(\theta_m))}{\sigma_m^2}} \tag{203}$$

When $\hat{\theta}_m$ is the maximum likelihood estimator of θ_m then $\hat{\sigma}^2 = \frac{1}{ST}(y - f(\hat{\theta}_m))^H(y - f(\hat{\theta}_m))$.Using this r_m reduces to

$$r_m = 2ST\ln\left(\frac{\hat{\sigma}_0^2}{\hat{\sigma}_{m-1}^2}\right) \tag{204}$$

where $\hat{\sigma}_0^2 = \frac{1}{ST}\sum_{s=1}^{S}\sum_{t=1}^{T} y(s,t)^H y(s,t)$ as $\Psi(S,T) \to \infty$. when $m = 1$, clearly $r_1 = 0$. Under normal assumptions of noise, M.L.E $\hat{\theta}_k$ of θ_k is same as the non- linear least sqaure estimate, we conclude using results from [6,5] as $\Psi(S,T) \to \infty$,

$$\hat{\alpha}_j \to \alpha_j \text{ for j=1,2,\ldots,m; } m \leq m_o \tag{205}$$

Case 1:$m \leq m_o$**(Underestimation)**
$\forall m \leq m_o$,

$$\hat{\sigma}_m^2 \to \sigma^2 + \sum_{j=m+1}^{m_o} \alpha_j^H \alpha_j \text{a.s. and } \hat{\sigma}_{m_o}^2 \to \sigma^2 \text{ a.s. as } \Psi(S,T) \to \infty \tag{206}$$

for $2 \leq m \leq m_o$ (207)

$$\frac{r_m}{ST} = 2\ln\left(\frac{\sigma^2 + \sum_{j=1}^{m_o} \alpha_j^H \alpha_j}{\sigma^2 + \sum_{j=m}^{m_o} \alpha_j^H \alpha_j}\right) \text{ a.s. as } \Psi(S,T) \to \infty \tag{208}$$

$$= 2\ln\left(1 + \frac{\sum_{j=1}^{m-1} \alpha_j^H \alpha_j}{\sigma^2 + \sum_{j=m}^{m_o} \alpha_j^H \alpha_j}\right) \text{ a.s. as } \Psi(S,T) \to \infty \tag{209}$$

since the R.H.S is bounded by definition of model and is strictly positive. Hence, we have $r_m = O(n)$ a.s. for all $m \leq m_o$.
Case 2: $m > m_o$**(Overestimation)**
For $m > m_o$, from Lemma 1 and lemma 2 we have

$$\frac{r_m}{ST} = 2\ln\left(\frac{\sigma^2 + \sum_{j=1}^{m_o} \alpha_j^H \alpha_j}{\sigma^2 + \frac{G_{m-m_o}}{ST} + o\left(\frac{(\ln(ST)\ln S \ln T)^{1/2}}{ST}\right)}\right) \text{ a.s. as } \Psi(S,T) \to \infty \tag{210}$$

(211)

where $G_k = \sum_{j=1}^{k} I_\epsilon(\beta_{m_o+j}, \omega_{m_o+j})$ $\forall k = 1, 2, \ldots, \tilde{m} - m_o$. Using (183), $G_k = O(\ln(ST))$

$$\frac{G_{m-m_o}}{ST} \to 0 \text{ a.s and } \ o\left(\frac{(\ln(ST)\ln S \ln T)^{1/2}}{ST}\right) \to 0 \text{ as } \Psi(S,T) \to \infty \tag{212}$$

Combining this we have for all $m > m_o$

$$\frac{r_m}{ST} = 2\ln\left(1 + \frac{\sum_{j=1}^{m_o} \alpha_j^H \alpha_j}{\sigma^2}\right) = O(1) \text{ a.s. as } \Psi(S,T) \to \infty \tag{213}$$

Hence, we get the desired result.
Lemma4: Under the assumptions A1-A3, ρ_m defined in (13) satisfies

$$\rho_m = \begin{cases} O(ST) & m \leq m_o \\ O_p(1) & m = m_o+1, m_o+2, \ldots, \tilde{m} \end{cases} \tag{214}$$

Proof:

$$\rho_m = 2\ln\left[\frac{f_{\tilde{m}}(y, \hat{\theta}^*_{\tilde{m}})}{f_{m-1}(y, \hat{\theta}^*_{m-1})}\right] \tag{215}$$

$$= 2ST\ln\left(\frac{\hat{\sigma}^2_{m-1}}{\hat{\sigma}_{\tilde{m}^2}}\right) \tag{216}$$

Using lemma 2

$$\hat{\sigma}^2_{\tilde{m}} = \hat{\sigma}^2_{m_o} - \sum_{j=1}^{\tilde{m}-m_o} \frac{I_\epsilon(\hat{\omega}_{m_o+j})}{ST} + o\left(\frac{(\ln(ST)\ln S \ln T)^{1/2}}{ST}\right) \text{ a.s. as } \Psi(S,T) \to \infty \tag{217}$$

Also, as $I_\epsilon(\beta, \omega) \geq 0 \forall \omega$

$$\frac{1}{ST}\sum_{j=1}^{\tilde{m}-m_o} I_\epsilon(\hat{\beta}_{m_o+j}, \hat{\omega}_{m_o+j}) \leq \frac{\tilde{m}-m_o}{ST} \sup_{(\beta,\omega)} I_\epsilon(\beta,\omega) \leq (\tilde{m}-m_o)\frac{\ln(ST)}{ST}\frac{\sup_{(\beta,\omega)} I_\epsilon(\beta,\omega)}{\ln(ST)} \to 0 \text{ a.s. as } \Psi(S,T) \to \infty \tag{218}$$

Case 1:$m \leq m_o$**(Underestimation)**
$\forall m \leq m_o$ as $\Psi(S,T) \to \infty$

$$\frac{\rho_m}{ST} \to 2\ln\left(\frac{\sigma^2 + \sum_{j=m}^{m_o} \alpha_j^H \alpha_j}{\sigma^2}\right) \text{ a.s} \tag{219}$$

The R.H.S is strictly positive and bounded and thus we have

$$\forall m \leq m_o \quad \rho_m = O(ST)\text{a.s. as } \ \Psi(S,T) \to \infty \tag{220}$$

Case 2:$m > m_o$**(overestimation)**
From asymptotic theory of likelihood ratios(see [8, 9, 10]) that as $\Psi(S,T) \to \infty$

$$\rho_m = 2ST\ln\left(\frac{\hat{\sigma}^2_{m-1}}{\hat{\sigma}_{\tilde{m}^2}}\right) \sim \chi^2_{4(\tilde{m}-m+1)} \tag{221}$$

where χ_k^2 denotes the chi square distribution with k degrees of freedom. Therefore, ρ_m is bounded in probability i.e. $O_p(1)$. Using the same we get the desired result.

Remark 2: Using lemma 3 and lemma 4 the ratio $\frac{\ln(r_m+1)}{\ln(\rho_m+1)}$ therefore satisfies

$$\frac{\ln(r_m+1)}{\ln(\rho_m+1)} = \begin{cases} O(1) & m \leq m_o \\ O(\ln(ST)) & m = m_o+1, m_o+2, \ldots, \tilde{m} \end{cases} \tag{222}$$

Theorem: Under A1-A3, if m_o is true order($m_o \leq \tilde{m}$) and $\hat{m}$ is the estimated model order using PAL rule then

$$P(\hat{m} \neq m_o) \rightarrow 0 \text{ as } \Psi(S,T) \rightarrow \infty \tag{223}$$

Proof: Under assumption of the model the PAL rule mentioned in (132) can also be written as

$$PAL(m) = 2ST\ln(\hat{\sigma}_m^2) + (4m+1)\ln(4\tilde{m}+1)\frac{\ln(r_m+1)}{\ln(\rho_m+1)} + \gamma \tag{224}$$

where γ is a constant.

Case 1: Underestimation($m \leq m_o$)

$$PAL(m) - PAL(m_o) = 2ST\ln(\frac{\hat{\sigma}_m^2}{\hat{\sigma}_{m_o}^2}) + (4m+1)\ln(4\tilde{m}+1)\frac{\ln(r_m+1)}{\ln(\rho_m+1)} - (4m_o+1)\ln(4\tilde{m}+1)\frac{\ln(r_{m_o}+1)}{\ln(\rho_{m_o}+1)} \tag{225}$$

Using lemma 1

- $\hat{\sigma}_{m_o}^2 \rightarrow \sigma^2$ a.s. as $\Psi(S,T) \rightarrow \infty$
- $\forall m \leq m_o,\ \hat{\sigma}_m^2 \rightarrow \sigma^2 + \sum_{j=m+1}^{m_o} \alpha_j^H \alpha_j$ a.s. as $\Psi(S,T) \rightarrow \infty$

$\forall m < m_o$

$$\ln(\frac{\hat{\sigma}_m^2}{\hat{\sigma}_{m_o}^2}) \rightarrow 2\ln\left(1 + \frac{\sum_{j=m+1}^{m_o} \alpha_j^H \alpha_j}{\sigma^2}\right) \text{ a.s as } \Psi(S,T) \rightarrow \infty \tag{226}$$

where R.H.S. is strictly positive and bounded by the definition of the model. Also, using lemma 3 and 4 we know that $r_m = O(ST)$ and $\rho_m = O(ST)$ a.s.

$$(4m+1)\ln(4\tilde{m}+1)\frac{\ln(r_m+1)}{\ln(\rho_m+1)} - (4m_o+1)\ln(4\tilde{m}+1)\frac{\ln(r_{m_o}+1)}{\ln(\rho_{m_o}+1)} \tag{227}$$

$$= (4m+1)\ln(4\tilde{m}+1)\frac{O(\ln n)}{O(\ln n)} - (4m_o+1)\ln(4\tilde{m}+1)\frac{O(\ln n)}{O(\ln n)} \text{ a.s as } \Psi(S,T) \rightarrow \infty \tag{228}$$

$$PAL(m) - PAL(m_o) = O(ST) + (4m+1)\ln(4\tilde{m}+1)\frac{O(\ln(ST))}{O(\ln(ST))} - (4m_o+1)\ln(4\tilde{m}+1)\frac{O(\ln(ST))}{O(\ln(ST))} \text{ a.s.} \tag{229}$$

$$\frac{1}{ST}(PAL(m) - PAL(m_o)) \rightarrow 2\ln\left(1 + \frac{\sum_{j=m+1}^{m_o} \alpha_j^H \alpha_j}{\sigma^2}\right) \text{ a.s.} \tag{230}$$

as $\Psi(S,T) \rightarrow \infty$ where R.H.S. is a strictly positive and bounded quantity.Thus,
$PAL(m) > PAL(m_o) \quad \forall m < m_o$ with probability 1

$$P(\hat{m} < m_o) = P(PAL(m) < PAL(m_o) \text{ for some } m < m_o) \tag{231}$$

$$= P(\bigcup_{m<m_o} PAL(m) < PAL(m_o)) \tag{232}$$

$$\leq \sum_{m=1}^{m_o-1} P\left(\frac{1}{ST}(PAL(m) - PAL(m_o) < 0)\right) \rightarrow 0 \tag{233}$$

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

Case 2: Overestimation($m > m_o$) We observe that

$$PAL(m) - PAL(m_o) = 2ST\ln(\frac{\hat{\sigma}^2_m}{\sigma^2_{m_o}}) + (4m+1)\ln(4\tilde{m}+1)\frac{\ln(r_m+1)}{\ln(\rho_m+1)} - (4m_o+1)\ln(4\tilde{m}+1)\frac{\ln(r_{m_o}+1)}{\ln(\rho_{m_o}+1)} \tag{234}$$

We know from Remark 1 that $\forall m < m_o$ and for all n

$$\frac{\ln(r_m+1)}{\ln(\rho_m+1)} \geq \frac{\ln(r_{m_o}+1)}{\ln(\rho_{m_o}+1)} \tag{235}$$

as ratio is an increasing function of m.Therefore we can write (234) as

$$\frac{1}{\ln(ST)}\Bigg(PAL(m) - PAL(m_o)\Bigg) \geq \frac{1}{\ln(ST)}\Bigg(2ST\ln(\frac{\hat{\sigma}^2_m}{\hat{\sigma}^2_{m_o}}) + 4(m-m_o)\ln(4\tilde{m}+1)\frac{\ln(r_m+1)}{\ln(\rho_m+1)}\Bigg) \quad \forall S,T \tag{236}$$

$$P\Bigg(\frac{1}{\ln(ST)}\Bigg(PAL(m) - PAL(m_o)\Bigg) > 0\Bigg) = P\Bigg(\frac{1}{\ln(ST)}\Bigg(2ST\ln(\frac{\hat{\sigma}^2_m}{\hat{\sigma}^2_{m_o}})\Bigg) + \frac{1}{\ln(ST)}\Bigg(4(m-m_o)\ln(4\tilde{m}+1)\frac{\ln(r_m+1)}{\ln(\rho_m+1)}\Bigg) > 0\Bigg) \tag{237}$$

It follows from the asymptotic theory of likelihood ratios (see [109]) that,

$$2ST\ln\Bigg(\frac{\hat{\sigma}^2_m}{\hat{\sigma}^2_{m_o}}\Bigg) \sim \chi^2_{4(m-m_o)} \tag{238}$$

Thus, $\frac{1}{\ln(ST)}\Bigg(2ST\ln\Bigg(\frac{\hat{\sigma}^2_m}{\hat{\sigma}^2_{m_o}}\Bigg)\Bigg) = o_p(1)$.Also, $\frac{1}{\ln(ST)}4(m-m_o)\ln(4m+1) > 0$ and the ratio $\frac{1}{\ln(ST)}\Bigg(\frac{\ln(r_m+1)}{\ln(\rho_m+1)}\Bigg) = o(1) \ \ \forall S,T$

$$P\Bigg(\frac{1}{\ln(ST)}\Bigg(2ST\ln(\frac{\hat{\sigma}^2_m}{\sigma^2_{m_o}})\Bigg) + \frac{1}{\ln(ST)}\Bigg(4(m-m_o)\ln(4\tilde{m}+1)\frac{\ln(r_m+1)}{\ln(\rho_m+1)}\Bigg) > 0\Bigg) \to 1 \tag{239}$$

Using (239), we have that

$$P(\hat{m} > m_o) = P(PAL(m) < PAL(m_o) \text{ for some } m > m_o) \tag{240}$$
$$= P(\bigcup_{m>m_o} PAL(m) < PAL(m_o)) \tag{241}$$
$$\leq \sum_{m=m_o+1}^{\tilde{m}} P\Bigg(\frac{1}{\ln ST}(PAL(m) - PAL(m_o) < 0)\Bigg) \to 0 \tag{242}$$

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

$$P(\hat{m} \neq m_o) \to 0 \text{ as } \ \Psi(S,T) \to \infty \tag{243}$$

4 Numerical examples

We consider the following 1-D cisoid model for simulation to compare performance of PAL method based estimator with other popular model order selection rules:

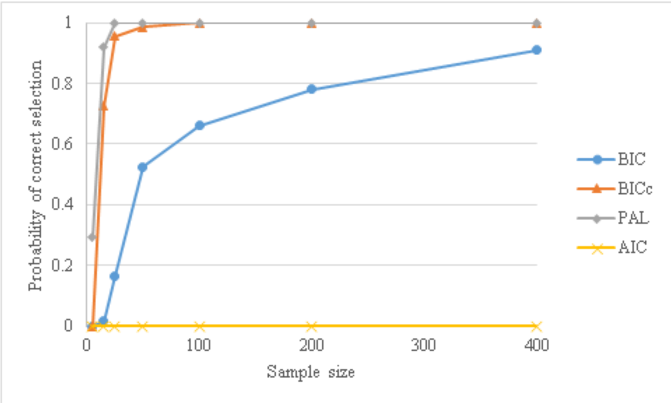
$$y_t = \sum_{k=1}^{m}\alpha_k e^{it\omega_k} + \epsilon_t$$

$$\alpha_1 = 3 + i2, \quad \alpha_2 = 2 + i1.66, \quad \alpha_3 = 1.75 + i$$
$$\omega_1 = 0.8\pi, \quad \omega_2 = 1.2\pi, \quad \omega_3 = 1.4\pi$$

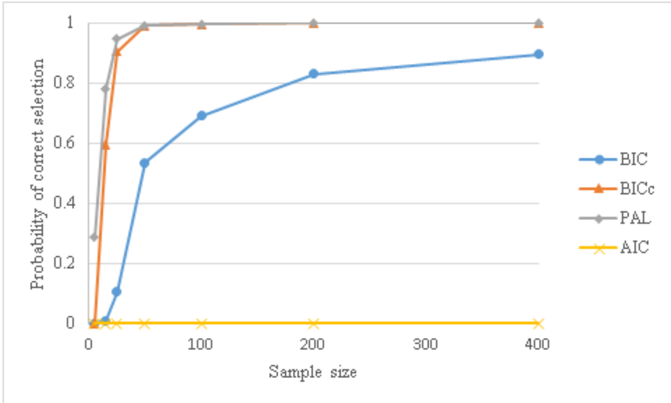
ϵ_t are i.i.d complex valued gaussian with zero mean s.t.

$$\epsilon_t = \epsilon_{t_R} + i\epsilon_{t_C} \tag{244}$$
$$\epsilon_{t_R} \sim \mathcal{N}(0, \sigma^2/2) \tag{245}$$
$$\epsilon_{t_C} \sim \mathcal{N}(0, \sigma^2/2) \tag{246}$$

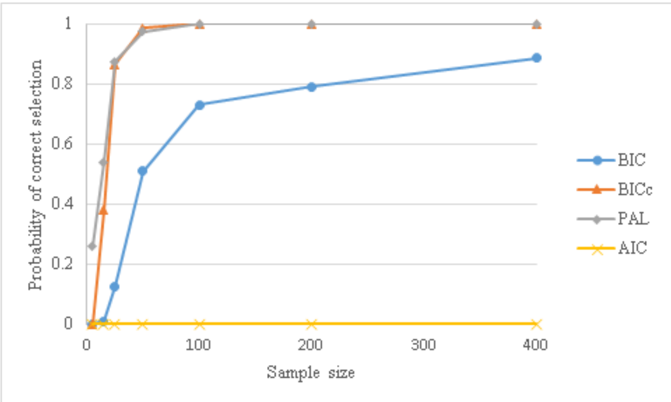
and ϵ_{t_R} and ϵ_{t_C} are independent and σ^2 varying from 1 to 30. We have considered the maximum model order to be 10, and sample size is varied from 5 to 200. We estimate the model order using different model order selection rules including PAL, BIC, BICc and AIC and report the probabilities of correct selection based on 200 simulation runs.



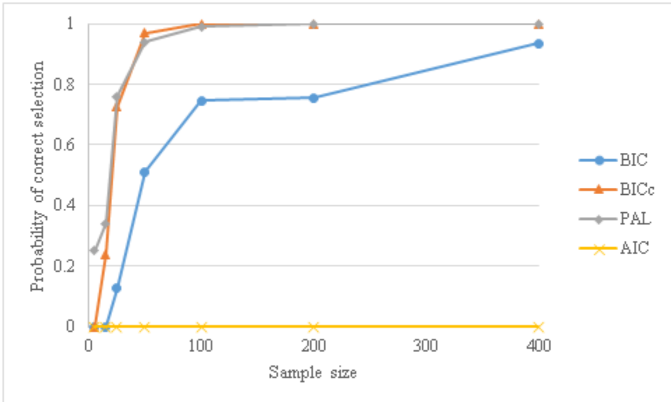
(a) $\sigma^2 = 1$



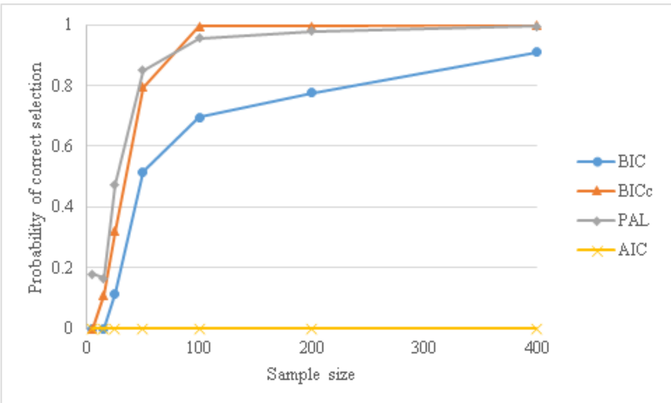
(b) $\sigma^2 = 3$



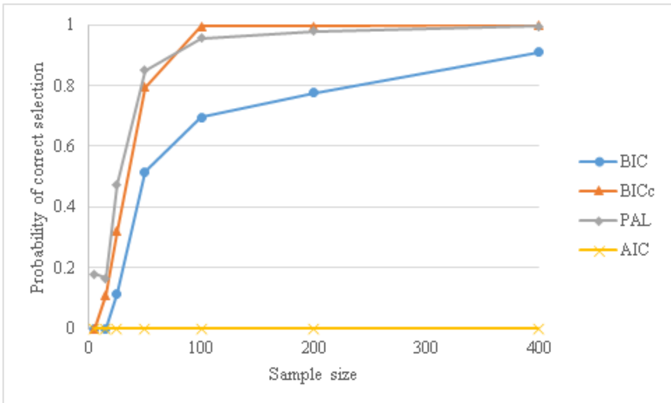
(c) $\sigma^2 = 6$



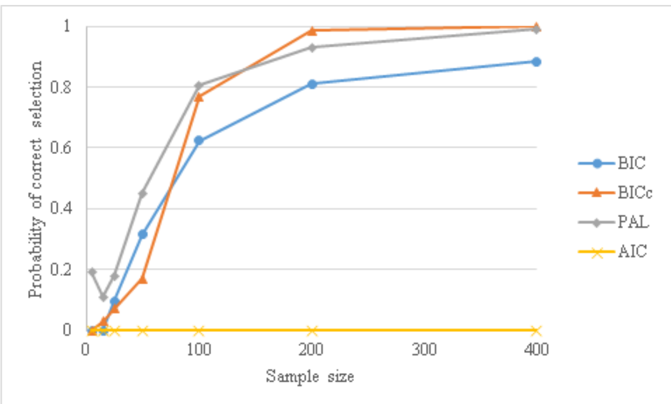
(d) $\sigma^2 = 9$



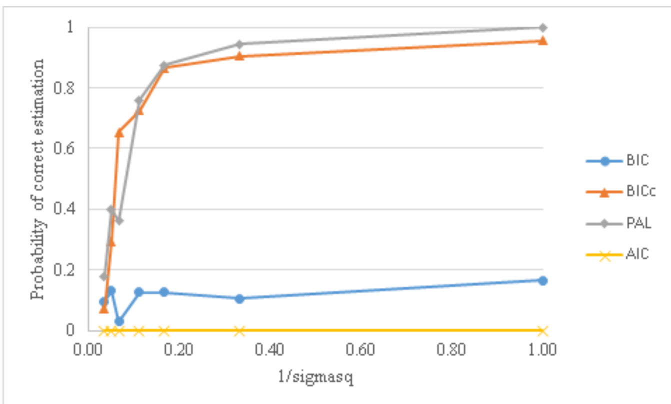
(e) $\sigma^2 = 15$



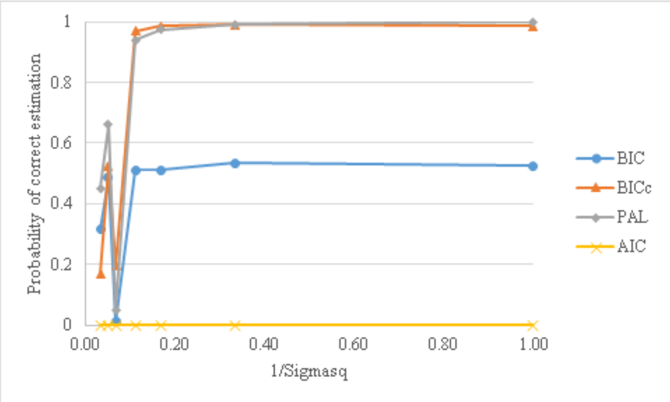
(f) $\sigma^2 = 20$



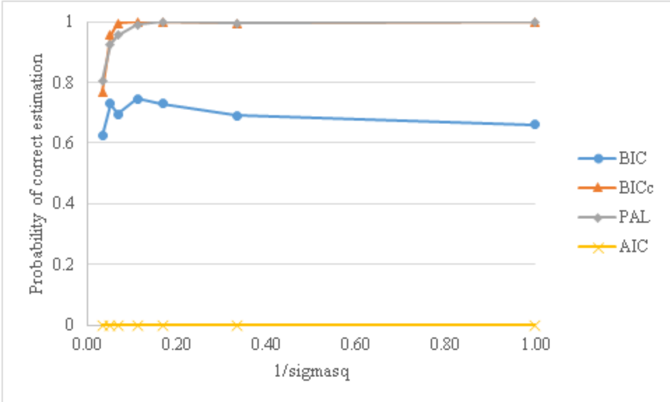
(g) $\sigma^2 = 30$



(h) $Samplesize = 25$



(a) $Samplesize = 50$



(b) $Samplesize = 100$

Figure 1: The following images shows plot of the probability of correct estimation of model order against different sample size for fixed values of variance of noise and plot the probability of correct estimation with different values of σ^2 for fixed sample sizes of 25,50 and 100.

5 Conclusion

In this project, we used the PAL based order selection method for estimating the number of cisoid in a 1-Dimensional and 2-Dimensional cisoid model and prove that the estimator based on PAL is asymptotically consistent. Using the simulation test that we conclude that PAL performs equally well or better than other popular order selection rules. We also concluded that as sample size increase or signal to noise ratio increases(with decrease in noise variance), the probability of correct estimation using PAL based approach goes to 1. Furthermore, we can now look at consistency of PAL based order selection rule for even more complex model like chirp signal to claim the superiority of PAL over other rules.

6 Reference

1. Chung, Kai Lai. A course in probability theory. Academic press, 2001.

2. Kundu, Debasis, and Amit Mitra. "Asymptotic properties of the least squares estimates of 2-D exponential signals." Multidimensional Systems and Signal Processing 7.2 (1996): 135-150.

3. D. Kundu and A. Mitra, "On Asymptotic Properties and Confidence Intervals for Exponential Signals ", Signal Processing , Vol. 72, No. 2, 129-139, 1999.

4. Francos, Joseph M., and Mark Kliger. "Strong consistency of the over-and under-determined LSE of 2-D exponentials in white noise." Proceedings.(ICASSP'05). IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005.. Vol. 4. IEEE, 2005.

5. Kliger, Mark, and Joseph M. Francos. "Strong consistency of a family of model order selection rules for estimating 2D sinusoids in noise." Statistics & Probability Letters 78.17 (2008): 3075-3081.

6. Prasad, Anurag, Debasis Kundu, and Amit Mitra. "Sequential estimation of the sum of sinusoidal model parameters." Journal of Statistical Planning and Inference 138.5 (2008): 1297-1313.

7. S.He, uniform convergence of weighted periodogram of stationary linear random fields, Chinese annals of mathematics 16B (1995) 331-340

8. Stoica, Petre, and Yngve Selen. "Model-order selection: a review of information criterion rules." IEEE Signal Processing Magazine 21.4 (2004): 36-47.

9. Wilks, Samuel S. "Sample criteria for testing equality of means, equality of variances, and equality of covariances in a normal multivariate distribution." The Annals of Mathematical Statistics (1946): 257-281.

10. Lehmann, Erich L., and Joseph P. Romano. "Testing Statistical Hypotheses (Springer Texts in Statistics)." (2005).

11. Stoica, Petre, and Prabhu Babu. "Model order estimation via penalizing adaptively the likelihood (PAL)." Signal processing 93.11 (2013): 2865-2871.

12. Surana, Khushboo, et al. "Estimating the order of sinusoidal models using the adaptively penalized likelihood approach: Large sample consistency properties." Signal Processing 128 (2016): 204-211.