MS Project

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

Anupreet Porwal[1]
Supervised by: Prof. Amit Mitra[1], Prof. Sharmishtha Mitra[1]

[1]Department of Mathematics and Statistics
Indian Institute of Technology Kanpur

November 8, 2016

Outline

1. Introduction
   - Model order selection
   - PAL rule
2. PAL consistency for 1-D cisoid model
   - 1-D cisoid model
   - PAL rule for 1-D cisoid
   - Consistency of PAL rule
   - Numerical examples
3. PAL rule consistency for 2-D Cisoid model
   - 2-Dimensional Cisoid Model
   - PAL rule for 2-D cisoid model
   - Consistency of PAL rule

Outline

1. Introduction
   - Model order selection
   - PAL rule
2. PAL consistency for 1-D cisoid model
   - 1-D cisoid model
   - PAL rule for 1-D cisoid
   - Consistency of PAL rule
   - Numerical examples
3. PAL rule consistency for 2-D Cisoid model
   - 2-Dimensional Cisoid Model
   - PAL rule for 2-D cisoid model
   - Consistency of PAL rule

Model order selection

- A fundamental problem in time series and signal processing
- Model order estimation determines the complexity of the model; allowing further estimation of other real valued model parameters

Model order selection

- A fundamental problem in time series and signal processing
- Model order estimation determines the complexity of the model; allowing further estimation of other real valued model parameters
- Two kind of estimation approaches:
  - maximum likelihood method(MLM) based rules: AIC, BIC, GIC etc.
  - Data driven approaches: bootstrapping, cross validation based approaches

Model order selection

- A fundamental problem in time series and signal processing
- Model order estimation determines the complexity of the model; allowing further estimation of other real valued model parameters
- Two kind of estimation approaches:
  - maximum likelihood method(MLM) based rules: AIC, BIC, GIC etc.
  - Data driven approaches: bootstrapping, cross validation based approaches
- Focus on MLM based order selection rules

Model order selection-II

- Information criterion rules like AIC[1] and BIC[2, 3] extensively studied
- AIC is not consistent and but is min-max optimal
- while for BIC, probability of correct order selection tends to 1 as $n \to \infty$ (consistent) but is not min-max optimal

Model order selection-II

- Information criterion rules like AIC[1] and BIC[2, 3] extensively studied
- AIC is not consistent and but is min-max optimal
- while for BIC, probability of correct order selection tends to 1 as $n \to \infty$ (consistent) but is not min-max optimal
- Ideally, the penalty term in order selection rules should follow
  1. for $m \leq m_o$, penalty should be small so that rule's value increases as m increases
  2. for $m > m_o$, penalty term should be large such that the rule's value decreases with increasing m

where m is model order and m_o is the true model order

PAL rule

- Introduced by Stoica and Babu[4], where model order estimation based on penalizing adaptively the likelihood
- Possess Oracle like properties that are mentioned in previous slide
- PAL based order estimation method shown consistent for non linear real sinusoidal model in [5]
- Here, we investigate the asymptotic statistical properties of the same for 1-D and 2-D cisoid models
- Simulation examples to compare PAL's performance with other order selection rules for 1-D cisoid models

Outline

1-Dimensional Cisoid model

- Building blocks of Digital signal processing

1-Dimensional Cisoid model

- Building blocks of Digital signal processing
- $\forall t = 1, 2, ..., n$

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

$$y_t = f(t, \theta_m) + \epsilon_t \qquad (2)$$

where
$\theta_m = (\alpha_{1_R}, \alpha_{1_C}, \omega_1, ..., \alpha_{m_R}, \alpha_{m_C}, \omega_m)'_{3m\mathrm{x}1}$=unknown signal parameter vector
α_{j_R} and α_{j_C} denotes the real and the imaginary part of α_j for j=1,..., m
m_o= true number of components in the the observed signal

1-Dimensional Cisoid model-Assumptions

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

$$\epsilon_t = \epsilon_{t_R} + i\epsilon_{t_C} \tag{3}$$
$$\epsilon_{t_R} \sim \mathcal{N}(0, \sigma^2/2) \tag{4}$$
$$\epsilon_{t_C} \sim \mathcal{N}(0, \sigma^2/2) \tag{5}$$

  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, α_k are bounded
- The true model parameter vector θ_{m_o} is an interior point in the parameter space $\Theta \subset \mathbb{R}^{3m_o}$

PAL rule for 1-D cisoid

Consider the two Generalized likelihood ratios,

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

where,
$\hat{\theta}^*_k = (\hat{\theta}_k, \hat{\sigma}^2_k) =$ M.L.E. of underlying signal and noise parameter vector θ^*_k
$f_0(y,\hat{\theta}^*_0) =$ p.d.f of y when M_0 is the model i.e., $f(\theta_0) = 0$

PAL rule for 1-D cisoid

Consider the two Generalized likelihood ratios,

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

where,
$\hat{\theta}^*_k = (\hat{\theta}_k, \hat{\sigma}^2_k) =$ M.L.E. of underlying signal and noise parameter vector θ^*_k
$f_0(y, \hat{\theta}^*_0) =$ 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:

$$PAL(m) = -2 \ln(f_m(y, \hat{\theta}^*_m)) + (3m+1) \ln(3\tilde{m}+1) \frac{\ln(r_m + 1)}{\ln(\rho_m + 1)} \tag{6}$$

$$\hat{m} =_{m \in \{1,2,\ldots,\tilde{m}\}} [PAL(m)] \tag{7}$$

where $\tilde{m}$ is the maximum number of cisoid components

Consistency of PAL rule

Lemma 1

Under the assumptions A1-A3, $\forall m \leq m_o$,as $n \to \infty$

$$\hat{\sigma}_m^2 = \sigma^2 + \sum_{j=1}^{m_o} \alpha_j^H \alpha_j - \sum_{j=1}^{m} \hat{\alpha}_j^H \hat{\alpha}_j + o(1)\text{a.s.} \tag{8}$$

Consistency of PAL rule

Lemma 1

Under the assumptions A1-A3, $\forall m \leq m_o$,as $n \to \infty$

$$\hat{\sigma}_m^2 = \sigma^2 + \sum_{j=1}^{m_o} \alpha_j^H \alpha_j - \sum_{j=1}^{m} \hat{\alpha}_j^H \hat{\alpha}_j + o(1)\text{a.s.} \tag{8}$$

Lemma 2

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

$$\hat{\sigma}_{m_o+k}^2 = \hat{\sigma}_{m_o}^2 - \sum_{j=1}^{k} \frac{I_\epsilon(\hat{\omega}_{m_o+j})}{n} + o\left(\frac{\ln n}{n}\right) \text{ a.s. as } n \to \infty \tag{9}$$

where $I_\epsilon(\omega)$ 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 $I_\epsilon(\omega)$

Consistency of PAL rule-II

Lemma 3

Under assumptions A1-A3, r_m satisfies

$$r_m = \begin{cases} 0 & m = 1 \\ O(n) & 2 \leq m \leq \tilde{m} \end{cases} \tag{10}$$

Consistency of PAL rule-II

Lemma 3

Under assumptions A1-A3, r_m satisfies

$$r_m = \begin{cases} 0 & m = 1 \\ O(n) & 2 \leq m \leq \tilde{m} \end{cases} \tag{10}$$

Lemma 4

Under the assumptions A1-A3, ρ_m satisfies

$$\rho_m = \begin{cases} O(n) & m \leq m_o \\ O_p(1) & m = m_o + 1, m_o + 2, \ldots, \tilde{m} \end{cases} \tag{11}$$

Consistency of PAL rule-III

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

$$P(\hat{m} \neq m_o) \rightarrow 0 \text{ as } \quad n \rightarrow \infty \tag{12}$$

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

$$P(\hat{m} \neq m_o) \to 0 \text{ as } \quad n \to \infty \tag{12}$$

Sketch of the Proof:Consider two cases:
Case I-**Underestimation** ($m \leq m_o$): Show that

$$\frac{1}{n}(PAL(m) - PAL(m_o)) \to 2\ln\left(1 + \frac{\sum_{j=m+1}^{m_o} \alpha_j^H \alpha_j}{\sigma^2}\right) \quad \text{a.s.} \tag{13}$$

as $n \to \infty$ where R.H.S. is a strictly positive and bounded quantity. Use it to show

$$P(\hat{m} < m_o) = P(PAL(m) < PAL(m_o) \text{ for some } m < m_o) \tag{14}$$

$$\leq \sum_{m=1}^{m_o-1} P\left(\frac{1}{n}(PAL(m) - PAL(m_o) < 0)\right) \to 0 \tag{15}$$

Sketch of proof- Continued

Case II-**Overestimation** ($m > m_o$): Using the asymptotic theory of likelihood ratios(see [6,7]) show that $\frac{1}{\ln(n)}\left(2n\ln\left(\frac{\hat{\sigma}_m^2}{\hat{\sigma}_{m_o}^2}\right)\right) = o_p(1)$

Use this to show:

$$P\left(\frac{1}{\ln n}\Big(PAL(m) - PAL(m_o)\Big) > 0\right) \to 1 \text{ as } n \to \infty \tag{16}$$

Thus,

$$P(\hat{m} > m_o) = P(PAL(m) < PAL(m_o) \text{ for some } m > m_o) \tag{17}$$

$$\leq \sum_{m=m_o+1}^{\tilde{m}} P\left(\frac{1}{\ln n}(PAL(m) - PAL(m_o) < 0)\right) \to 0 \tag{18}$$

as $n \to \infty$. This implies

$$P(\hat{m} \neq m_o) \to 0 \text{ as } \; n \to \infty \tag{19}$$

Numerical examples

Example

Simulation example

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

$$\alpha_1 = 3 + i2, \quad \alpha_2 = 2 + i1.66, \quad \alpha_3 = 1.75 + i$$

$$\omega_1 = 0.8\pi, \quad \omega_2 = 1.2\pi, \quad \omega_3 = 1.4\pi$$

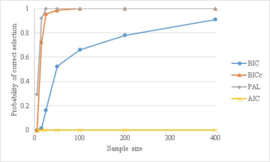
ϵ_t are i.i.d complex valued gaussian with zero mean s.t.

$$\epsilon_t = \epsilon_{t_R} + i\epsilon_{t_C} \tag{20}$$

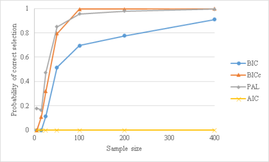
$$\epsilon_{t_R} \sim \mathcal{N}(0, \sigma^2/2) \tag{21}$$

$$\epsilon_{t_C} \sim \mathcal{N}(0, \sigma^2/2) \tag{22}$$

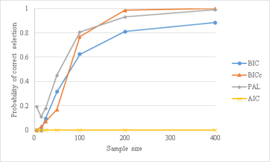
ϵ_{t_R} and ϵ_{t_C} are independent



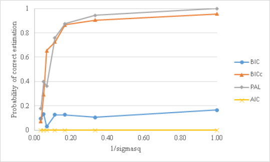
(a) $\sigma^2 = 1$



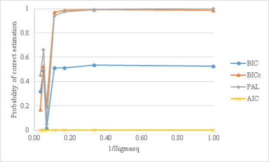
(b) $\sigma^2 = 15$



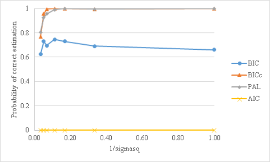
(c) $\sigma^2 = 30$



(d) $Samplesize = 25$



(e) $Samplesize = 50$



(f) $Samplesize = 100$

Figure: (a)-(c):Plot of the probability of correct estimation of model order against different sample size for fixed values of variance of noise; (d)-(f) plot the probability of correct estimation with different values of σ^2 for fixed sample sizes of 25,50 and 100

Outline

2-Dimensional Cisoid Model

- Widespread application in texture analysis

2-Dimensional Cisoid Model

- Widespread application in texture analysis
- $\forall s, t$ such that$1 \leq s \leq S; 1 \leq t \leq T$

$$y(s,t) = \sum_{k=1}^{m} \alpha_k e^{i(s\beta_k + t\omega_k)} + \epsilon(s,t) \quad (23)$$

$$y(s,t) = f(s,t,\theta_m) + \epsilon(s,t) \quad (24)$$

$\theta_m = (\alpha_{1_R}, \alpha_{1_C}, \beta_1, \omega_1, ..., \alpha_{m_R}, \alpha_{m_C}, \beta_m, \omega_m)'_{4m \times 1}$: vector of unknown signal parameters
m_o=True number of components in the the observed signal.

2-Dimensional Cisoid Model-Assumptions

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

$$\epsilon(s,t) = \epsilon_R(s,t) + i\epsilon_C(s,t) \tag{25}$$
$$\epsilon_R(s,t) \sim \mathcal{N}(0, \sigma^2/2) \tag{26}$$
$$\epsilon_C(s,t) \sim \mathcal{N}(0, \sigma^2/2) \tag{27}$$

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

- $\forall k = 1, 2, ..., m_o : (\beta_k, \omega_k) \in (0, 2\pi)\text{x}(0, 2\pi)$;
  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 $\Theta \subset \mathbb{R}^{4m_o}$

PAL rule for 2-D cisoid model

Using the two Generalized likelihood ratios defined before, the PAL rule for this model can then be defined as follows:

PAL rule for 2-D cisoid model

Using the two Generalized likelihood ratios defined before, the PAL rule for this model can then be defined as follows:

$$PAL(m) = -2\ln(f_m(y, \hat{\theta}_m^*)) + (4m+1)\ln(4\tilde{m}+1)\frac{\ln(r_m+1)}{\ln(\rho_m+1)} \tag{28}$$

$$\hat{m} =_{m \in \{1,2,\ldots,\tilde{m}\}} [PAL(m)] \tag{29}$$

where $\tilde{m}$ is the maximum number of components in the model

PAL rule for 2-D cisoid model

Using the two Generalized likelihood ratios defined before, the PAL rule for this model can then be defined as follows:

$$PAL(m) = -2\ln(f_m(y, \hat{\theta}^*_m)) + (4m+1)\ln(4\tilde{m}+1)\frac{\ln(r_m+1)}{\ln(\rho_m+1)} \tag{28}$$

$$\hat{m} =_{m \in \{1,2,\ldots,\tilde{m}\}} [PAL(m)] \tag{29}$$

where $\tilde{m}$ is the maximum number of components in the model
Remark:Let $\{\Psi_i\}$ be a sequence of rectangles s.t.

$$\Psi = \{(s,t) \in \mathbb{Z}^2 | 1 \leq s \leq S_i, 1 \leq t \leq T_i\} \tag{30}$$

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

$$\lim_{i \to \infty} min(S_i, T_i) \to \infty \text{ and } 0 < \lim_{i \to \infty} \left(\frac{S_i}{T_i}\right) < \infty \tag{31}$$

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

Consistency of PAL rule

Lemma 1

Under the assumptions A1-A3, $\forall m \leq m_o$,as $\Psi(S, T) \to \infty$

$$\hat{\sigma}_m^2 = \sigma^2 + \sum_{j=1}^{m_o} \alpha_j^H \alpha_j - \sum_{j=1}^{m} \hat{\alpha}_j^H \hat{\alpha}_j + o(1) \text{ a.s.} \tag{32}$$

Consistency of PAL rule

Lemma 1

Under the assumptions A1-A3, $\forall m \le m_o$,as $\Psi(S,T) \to \infty$

$$\hat{\sigma}_m^2 = \sigma^2 + \sum_{j=1}^{m_o} \alpha_j^H \alpha_j - \sum_{j=1}^{m} \hat{\alpha}_j^H \hat{\alpha}_j + o(1) \text{ a.s.} \tag{32}$$

Lemma 2

Under A1-A3, for any integer $k \ge 1$,as $\Psi(S,T) \to \infty$

$$\hat{\sigma}_{m_o+k}^2 = \hat{\sigma}_{m_o}^2 - \sum_{j=1}^{k} \frac{I_\epsilon(\hat{\beta}_{m_o+j}, \hat{\omega}_{m_o+j})}{ST} + o\left(\frac{(\ln(ST)\ln S \ln T)^{1/2}}{ST}\right) \text{ a.s.} \tag{33}$$

where $I_\epsilon(\beta,\omega)$ 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 $I_\epsilon(\beta,\omega)$

Consistency of PAL rule-II

Lemma 3

Under assumptions A1-A3, r_m satisfies

$$r_m = \begin{cases} 0 & m = 1 \\ O(ST) & 2 \leq m \leq \tilde{m} \end{cases} \tag{34}$$

Consistency of PAL rule-II

Lemma 3

Under assumptions A1-A3, r_m satisfies

$$r_m = \begin{cases} 0 & m = 1 \\ O(ST) & 2 \leq m \leq \tilde{m} \end{cases} \tag{34}$$

Lemma 4

Under the assumptions A1-A3, ρ_m satisfies

$$\rho_m = \begin{cases} O(ST) & m \leq m_o \\ O_p(1) & m = m_o + 1, m_o + 2, \ldots, \tilde{m} \end{cases} \tag{35}$$

Consistency of PAL rule-III

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

$$P(\hat{m} \neq m_o) \rightarrow 0 \text{ as } \Psi(S, T) \rightarrow \infty \tag{36}$$

Consistency of PAL rule-III

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

$$P(\hat{m} \neq m_o) \to 0 \text{ as } \Psi(S,T) \to \infty \tag{36}$$

Sketch of the Proof:Consider two cases:
Case I-**Underestimation** ($m \leq m_o$): Show that

$$\frac{1}{ST}(PAL(m) - PAL(m_o)) \to 2\ln\left(1 + \frac{\sum_{j=m+1}^{m_o} \alpha_j^H \alpha_j}{\sigma^2}\right) \text{ a.s.} \tag{37}$$

as $\Psi(S,T) \to \infty$ where R.H.S. is a strictly positive and bounded quantity. Use it to show

$$P(\hat{m} < m_o) = P(PAL(m) < PAL(m_o) \text{ for some } m < m_o) \tag{38}$$

$$\leq \sum_{m=1}^{m_o-1} P\left(\frac{1}{ST}(PAL(m) - PAL(m_o) < 0)\right) \to 0 \tag{39}$$

Sketch of proof- Continued

Case II-**Overestimation** ($m > m_o$): Using the asymptotic theory of likelihood ratios(see [6,7]) show that $\frac{1}{\ln(ST)}\left(2ST\ln\left(\frac{\hat{\sigma}_m^2}{\hat{\sigma}_{m_o}^2}\right)\right) = o_p(1)$

Use this to show:

$$P\left(\frac{1}{\ln(ST)}\Big(PAL(m) - PAL(m_o)\Big) > 0\right) \to 1 \text{ as } \Psi(S,T) \to \infty \tag{40}$$

Thus,

$$P(\hat{m} > m_o) = P(PAL(m) < PAL(m_o) \text{ for some } m > m_o) \tag{41}$$

$$\leq \sum_{m=m_o+1}^{\tilde{m}} P\left(\frac{1}{\ln(ST)}(PAL(m) - PAL(m_o) < 0)\right) \to 0 \tag{42}$$

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

$$P(\hat{m} \neq m_o) \to 0 \text{ as } \Psi(S,T) \to \infty \tag{43}$$

Conclusion

- The model order estimator based on PAL is asymptotically consistent for 1-D and 2-D cisoid models

Conclusion

- The model order estimator based on PAL is asymptotically consistent for 1-D and 2-D cisoid models
- Simulation tests suggests PAL performs equally well or better than other popular order selection rules

Conclusion

- The model order estimator based on PAL is asymptotically consistent for 1-D and 2-D cisoid models
- Simulation tests suggests PAL performs equally well or better than other popular order selection rules
- 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

Conclusion

- The model order estimator based on PAL is asymptotically consistent for 1-D and 2-D cisoid models
- Simulation tests suggests PAL performs equally well or better than other popular order selection rules
- 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
- Outlook
    - PAL consistency for chirp signal model
    - High SNR consistency of PAL based order selection rule

References

1. Akaike, Hirotugu. "A new look at the statistical model identification." IEEE transactions on automatic control 19.6 (1974): 716-723.
2. Schwarz, Gideon. "Estimating the dimension of a model." The annals of statistics 6.2 (1978): 461-464. Second item.
3. Rissanen, Jorma. "Modeling by shortest data description." Automatica 14.5 (1978): 465-471.
4. Stoica, Petre, and Prabhu Babu. "Model order estimation via penalizing adaptively the likelihood (PAL)." Signal processing 93.11 (2013): 2865-2871.

5. Surana, K., Mitra, S., Mitra, A., and Stoica, P. (2016). Estimating the order of sinusoidal models using the adaptively penalized likelihood approach: Large sample consistency properties. Signal Processing, 128, 204-211.
6. 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.
7. Lehmann, Erich L., and Joseph P. Romano. "Testing Statistical Hypotheses (Springer Texts in Statistics)." (2005).