STOCHASTIC COMPLEXITY FOR THE ESTIMATION OF SINE-WAVES IN COLORED NOISE

Ciprian Doru Giurcăneanu

Tampere University of Technology, Institute of Signal Processing
P.O. Box 553, FIN-33101 Tampere, Finland
ciprian.giurcaneanu@tut.fi

ABSTRACT

During recent years the advances in stochastic complexity (SC) have led to new exact formulae or to sharper approximations for large classes of models. We focus on the use of the SC to estimate the structure for the model of sine-waves in Gaussian autoregressive noise. Since the evaluation of SC relies on the determinant of the Fisher information matrix (FIM), the computation of FIM is revisited. It is shown for small and moderate sample sizes that SC compares favorably with other well-known criteria such as: BIC, KICc and GAIC.

Index Terms— Minimum Description Length principle, sinusoidal regression, autoregressive processes, structure selection, Fisher information matrix.

1. INTRODUCTION AND PRELIMINARIES

We address the estimation of the number of sinusoidals observed in additive noise with unknown correlation structure. To formulate the problem, we consider the data model

$$y_t = x_t + e_t,\ t \in \{0, \ldots, N-1\},$$
$$x_t = \sum_{k=1}^{K} \alpha_k \cos(\omega_k t + \phi_k), \quad (1)$$

where y_t denotes the measurements, x_t is the noise-free signal and e_t is the colored Gaussian noise.

To ensure the identifiability of the parameters, we assume as usual that the amplitudes α_k are strict positive and the frequencies ω_k belong to the interval $(0, \pi)$. The frequencies are distinct and, without loss of generality, $\omega_1 < \cdots < \omega_K$. Both the amplitudes and the frequencies are non-random parameters that will be estimated from the available measurements.

Two different hypotheses will be considered for modeling the phases $\phi_k \in [-\pi, \pi)$: H_{dp} - the phases are unknown deterministic constants; H_{rp} - the phases are independent and uniformly distributed random variables that are also independent of e_t. For both assumptions, the statistical properties of y_t have been investigated in previous studies, and more details can be found, for example, in [1].

In line with the approach from [1][2][3] and the references therein, we model the noise e_t as a stable autoregressive (AR) process with order M:

$$e_t = -\sum_{m=1}^{M} a_m e_{t-m} + w_t, \quad (2)$$

This work was supported by the Academy of Finland, project No. 113572 and 213462.

where w_t is a sequence of independent and identically distributed Gaussian random variables with zero mean and variance τ. Since we consider only the case of real sinusoidals in real AR noise, we emphasize that the white random process w_t and the coefficients a_m, $1 \leq m \leq M$, are real-valued.

Recall the definition of the local SNR for the k-th sinusoid, $\text{SNR}_k = \frac{\alpha_k^2/2}{\tau/|A(\omega_k)|^2}$, where $A(\omega_k) = 1 + \sum_{m=1}^{M} a_m \exp(-jm\omega_k)$ and $j = \sqrt{-1}$ [4].

Based on (1) and (2), we note that the parameters of the model are τ and $\theta = (\xi, \mathfrak{a})$, where $\xi = (\xi_1, \ldots, \xi_K)$ with the convention $\xi_k = (\alpha_k, \omega_k, \phi_k)$ for the k-th sine-wave. The notation $\mathfrak{a}$ is employed for the set of the AR coefficients $(a_1, \ldots, a_M)$.

Because the model structure $\gamma = (K, M)$ is not known a priori, we resort to the traditional model selection procedure that comprises two steps: (a) for all pairs of integers $\gamma = (K, M)$ that satisfy $0 \leq K \leq K_{max}$ and $0 \leq M \leq M_{max}$, estimate the model parameters $\hat{\theta}_\gamma$ from the observations $y^N = y_0, \ldots, y_{N-1}$; (b) evaluate an information theoretic criterion for all γs considered at the first step, and choose the model structure $\hat{\gamma}$ that minimizes the criterion.

The most popular information theoretic rules for order selection can be reduced to a common form with two terms, where the first one is given by the minus maximum log-likelihood. The second term is a penalty coefficient that depends on the number of parameters of the model and, for some criteria, also on the sample size N [5]. In general, the selection rules used in practical applications are derived for $N \to \infty$, and the asymptotic approximations could potentially yield false conclusions when the sample size is small or moderate.

During recent years, the advances in stochastic complexity (SC) have led to new exact formulae or to sharper approximations for large classes of models [6][7][8], but the use of the new results in signal processing is scarce. We illustrate next how SC can be applied to estimate the structure for the model of sine-waves in Gaussian AR noise.

The rest of the paper is organized as follows. Section 2 is focused on the SC formula for sinusoids in AR noise. Because the evaluation of SC relies on the determinant of the Fisher information matrix (FIM), the computation of FIM is revisited in Section 3. SC and three other well-known criteria are compared in Section 4 to evaluate their performances in estimating the number of sinusoids from simulated data.

2. SC FOR SINE-WAVES IN AR NOISE

Under mild assumptions on the maximum likelihood (ML) estimates $\hat{\theta} = \hat{\theta}(y^N)$ and $\hat{\tau} = \hat{\tau}(y^N)$, we can use the expression of SC given in [6]. We employ the notation $f(y^N;\theta,\tau)$ for the likelihood function of y^N, thus $\ln f(y^N;\hat{\theta},\hat{\tau})$ is the maximum log-likelihood. Θ denotes the parameter space and $\mathbf{J}_N(\theta) = E\left[-\frac{\partial^2 \ln f(y^N;\theta,\tau)}{\partial\theta\partial\theta^\top}\right]$ is the Fisher information matrix (FIM).

It is well-known that the SC derivation is grounded in information theory. The predictive approach treats the white noise variance τ as a nuisance parameter in the sense that the code length to describe it is not included in the SC formula. In our application, we noticed that considering the cost for transmitting τ in the evaluation of the code length has a small influence on the estimation results, and for brevity we discuss in this study only the predictive approach.

To apply the SC formula from [6] we have to compute the term $\ln\int_\Theta |\mathbf{J}_N(\theta)|^{1/2}\mathrm{d}\theta$ for which it is difficult to find a closed-form expression. For the simpler AR case, the integral was evaluated with Monte Carlo techniques in [9]. In our application, a supplementary difficulty is due to the amplitudes α_k, for which the domain does not have an upper bound. This makes us to prefer the SC formula derived in [8]:

$$\mathrm{SC}(y^N;K,M) = -\ln f(y^N;\hat{\theta},\hat{\tau}) + \ln|\mathbf{J}_N(\hat{\theta})|^{1/2} + \sum_{i=1}^{3K+M}\ln(|\hat{\theta}_i| + N^{-1/4}) \quad (3)$$

We check in Appendix how the conditions for the derivations from [8] are fulfilled for the model of sinusoids in Gaussian AR noise. In Appendix we also give more details on the accuracy of the approximation in formula (3).

To gain more insight, we briefly analyze what SC formula (3) becomes for $N\to\infty$. Based on the results from the Appendix, we get the well-known asymptotic identity $\lim_{N\to\infty}\ln|\mathbf{J}_N(\hat{\theta})|^{1/2} = \frac{5K+M}{2}\ln N$, and it is easy to notice that the sum of the first two terms in SC is equivalent with the Bayesian information criterion (BIC):

$$\mathrm{BIC}(y^N;K,M) = -\ln f(y^N;\hat{\theta},\hat{\tau}) + \frac{5K+M}{2}\ln N. \quad (4)$$

More details on the derivation of BIC can be found in [5] and the references therein. The formula (4) was also obtained in [2] as a crude version of the Minimum Description Length criterion. In [10], the use of the *maximum a posteriori probability* methodology in conjunction with asymptotic approximations led also to (4) for the particular case of white noise.

We focus on the last term in (3), and for simplicity we assume $M = 0$ (white noise). The two-step encoding procedure adopted in [8] employs first a uniform quantization of Θ that is performed with the same precision for all the parameters. The term $N^{-1/4}$ in (3) is due to the option from [8] to select this precision based on the minimum eigenvalue of FIM. If zero does not belong to the domain of the parameter θ_i, then $\hat{\theta}_i \neq 0$ and $\ln(|\hat{\theta}_i| + N^{-1/4})$ is much smaller than $\frac{1}{2}\ln N$ [8]. Hence the term $\ln(|\hat{\theta}_i| + N^{-1/4})$ becomes important only when $\hat{\theta}_i = 0$. Since among the ξ parameters only the phases can be equal to zero, asymptotically the penalty term in SC formula takes values between $\frac{9K}{4}\ln N$ and $\frac{5K}{2}\ln N$. Based on this observation, it is easy to show that the necessary conditions for consistency from [11] are verified.

3. COMPUTATIONAL ISSUES

The use of (3) is very appealing from computational viewpoint, but it was already pointed out in [8] that (3) is not invariant under reparametrization. Due to this reason, we prefer to use as parameters for the AR noise the magnitudes and angles of the poles instead of the coefficients.

SC	Hypothesis	$\mathbf{J}_N(\xi)$	$\mathbf{J}_N(\mathfrak{g})$
SCp	Hrp	exact	exact
SCa	Hrp/H$_{\mathrm{dp}}$	asymptotic	asymptotic
SCe	H$_{\mathrm{dp}}$	exact	exact

Table 1. Nomenclature for SC when various formulae for FIM are used in calculations.

More precisely, let us assume that the poles of the AR noise are $g_1,\ldots,g_M$: if the poles $g_1,\ldots,g_{M_1}$ are real-valued, then the pure complex poles $g_{M_1+1},\ldots,g_M$ occur in complex conjugate pairs because the coefficients $\mathfrak{a}$ are real-valued. Instead of $\theta = (\xi,\mathfrak{a})$, we will apply the parametrization $\eta = (\xi,\mathfrak{g})$, where $\mathfrak{g} = (g_1,\ldots,g_{M_1},|g_{M_1+1}|,\psi_{g_{M_1+1}},\ldots,|g_{M-1})|,\psi_{g_{M-1}})$ and for a complex pole g_i, the symbol ψ_{g_i} denotes its angle. Remark the range of the entries of $\mathfrak{g}$: we have $g_i \in (-1,1)$ for $1 \le i \le M_1$, and for the rest of the parameters $|g_i| \in (0,1)$ and $\psi_{g_i} \in (0,\pi)$.

In [1], it is proven that $|\mathbf{J}_N(\eta)|$ can be factorized as $|\mathbf{J}_N(\xi)| \times |\mathbf{J}_N(\mathfrak{g})|$, where the block $\mathbf{J}_N(\xi)$ corresponds to the signal parameters, and the block $\mathbf{J}_N(\mathfrak{g})$ corresponds to the noise parameters. The reference [1] gives also fast algorithms for the evaluation of $\mathbf{J}_N(\xi)$. $\mathbf{J}_N(\mathfrak{g})$ has the same expression as in the pure AR case, and for its calculation we resort to the exact and asymptotic formulae from [12]. We note that the formulae are given for the $\mathfrak{a}$ parameters and their change to $\mathfrak{g}$ parameters is straightforward [13].

Applying the exact or asymptotic formulae for $\mathbf{J}_N(\xi)$ and $\mathbf{J}_N(\mathfrak{g})$ leads to various expressions for SC. In Table 1, we explain the nomenclature for SC when FIM in (3) is evaluated with various formulae.

4. EXPERIMENTAL RESULTS

In all the examples presented next, we resort to the RELAX algorithm that performs a decoupled parameter estimation for the sinusoids and the AR noise [3]. In our simulations we have used for the implementation of RELAX Matlab functions that are publicly available at *http://www.uni-kassel.de/fb16/hfk/neu/toolbox*.

Asymptotically both RELAX and ML yield statistically efficient estimates, and the use of RELAX is recommended due to its lower computational burden [3][4].

For $\gamma = (K,M)$, let $\hat{\xi}_k$ be the parameters of the k-th sinusoid estimated with RELAX. We denote $\hat{e}_t = y_t - \sum_{k=1}^K \hat{\alpha}_k\cos(\hat{\omega}_k t + \hat{\phi}_k)$, and let $\hat{\mathfrak{a}}$ be the coefficients of the AR noise determined from the sequence $\hat{e}_0,\ldots,\hat{e}_{N-1}$. We further define the residual sum of squares as $\mathrm{RSS}_\gamma = \sum_{t=0}^{N-1}\left[\hat{e}_t + \sum_{m=1}^M \hat{a}_m\hat{e}_{t-m}\right]^2$, with the convention that $\hat{e}_t = 0$ for $t < 0$.

In SC formula (3) and the BIC formula (4), $-\ln f(y^N;\hat{\theta},\hat{\tau})$ is evaluated as $\frac{N}{2}\ln\mathrm{RSS}_\gamma$ after discarding the terms that do not depend on γ. The performances of SC and BIC are compared in our simulations with two other criteria: KICc and GAIC. KICc was derived in [14] as a unbiased Kullback Information Criterion for linear regression models with i.i.d Gaussian noise. Since then its application was extended also to other classes of models, see for example [15] and the references therein. GAIC is a generalized Akaike Information Criterion that was traditionally used in conjunction with the RELAX algorithm [3].

Example 1 ($K = 2$)									Example 2 ($K = 2$)								
SNR$_2$=-3.00 dB									SNR$_2$=1.00 dB								
N	30	40	50	60	70	80	90	100	N	30	40	50	60	70	80	90	100
SCp	37	69	76	**85**	**98**	**98**	**98**	95	SCp	**36**	**52**	**73**	**73**	76	85	85	94
SCa	36	68	73	84	95	95	97	95	SCa	34	45	71	68	75	84	85	94
SCe	31	62	64	77	89	85	90	84	SCe	27	28	68	**73**	74	85	85	94
BIC	33	50	58	66	83	90	93	**96**	BIC	6	28	67	67	**78**	**89**	**89**	**97**
KICc	**61**	**74**	**77**	48	52	51	37	40	KICc	21	23	35	36	32	39	35	44
GAIC	3	1	19	23	44	62	68	80	GAIC	6	8	32	43	61	74	84	**97**
SNR$_2$=-1.00 dB									SNR$_2$=3.00 dB								
N	30	35	40	45	50	60	80	100	N	30	35	40	45	50	60	80	100
SCp	72	86	**91**	**95**	**97**	**98**	**99**	99	SCp	**57**	**54**	**74**	**83**	**94**	**89**	95	96
SCa	72	83	84	90	91	**98**	97	98	SCa	56	52	63	75	90	**89**	94	96
SCe	58	64	76	77	77	86	89	92	SCe	48	42	17	29	87	85	94	96
BIC	50	65	72	81	82	89	97	99	BIC	31	32	44	59	84	84	**97**	**98**
KICc	**83**	**89**	81	79	74	74	48	44	KICc	38	39	30	38	54	35	44	43
GAIC	5	23	32	48	61	75	88	**100**	GAIC	11	7	15	21	51	66	96	**98**
SNR$_2$=0.00 dB									SNR$_2$=5.00 dB								
N	25	30	35	40	45	50	75	100	N	25	30	35	40	45	50	75	100
SCp	78	**82**	90	**93**	**96**	**98**	**98**	97	SCp	**49**	**63**	**72**	**76**	**88**	**88**	93	98
SCa	70	80	78	85	84	90	95	97	SCa	47	62	68	61	74	87	92	98
SCe	56	67	72	74	75	73	87	89	SCe	38	54	45	14	15	83	93	98
BIC	52	54	70	75	87	88	93	**100**	BIC	23	25	51	47	81	83	98	**100**
KICc	**79**	81	**91**	87	81	76	55	33	KICc	38	32	44	30	42	50	49	43
GAIC	10	24	36	46	65	79	**98**	**100**	GAIC	13	18	27	31	62	77	**100**	**100**
Example 3 ($K = 2$)									Example 4 ($K = 3$)								
SNR$_1$=1.00 dB									SNR$_1$=-5.00 dB								
N	30	40	50	60	70	80	90	100	N	30	40	50	60	70	80	90	100
SCp	**74**	**74**	**88**	**89**	**80**	**79**	**81**	**91**	SCp	21	**75**	76	**71**	76	**76**	**85**	**85**
SCa	71	65	81	85	77	77	**81**	87	SCa	21	**75**	**77**	**71**	**77**	**76**	**85**	84
SCe	64	69	80	**89**	**80**	**79**	80	**91**	SCe	35	66	72	68	72	**76**	84	83
BIC	42	56	70	77	73	73	72	82	BIC	19	60	72	67	73	67	79	76
KICc	60	68	70	61	58	51	48	51	KICc	**52**	68	61	56	56	47	53	39
GAIC	13	41	63	68	74	71	72	80	GAIC	0	7	25	44	65	62	72	71
SNR$_1$=3.00									SNR$_1$=-3.00 dB								
N	30	35	40	45	50	60	80	100	N	30	40	50	60	70	80	90	100
SCp	**84**	**89**	**93**	**82**	**88**	87	**85**	82	SCp	35	85	**82**	**91**	**88**	**90**	**96**	**90**
SCa	74	82	85	74	85	**88**	82	80	SCa	35	**86**	**82**	**91**	**88**	89	**96**	88
SCe	72	77	81	76	85	86	**85**	82	SCe	52	78	75	85	84	85	93	86
BIC	49	64	77	68	77	82	79	87	BIC	27	80	77	83	84	83	92	**90**
KICc	74	79	83	72	71	69	59	40	KICc	**69**	84	71	71	58	56	55	48
GAIC	49	52	74	78	83	86	80	**91**	GAIC	0	29	57	61	72	65	80	82
SNR$_1$=5.00 dB									SNR$_1$=-1.00 dB								
N	25	30	35	40	45	50	75	100	N	30	35	40	45	50	60	80	100
SCp	**90**	**93**	**90**	**91**	**97**	**95**	90	90	SCp	41	**85**	**93**	**89**	**93**	**93**	94	86
SCa	85	87	84	77	81	92	86	90	SCa	40	**85**	92	**89**	91	91	90	82
SCe	79	78	74	83	90	93	88	88	SCe	50	65	86	71	79	82	86	85
BIC	52	60	67	77	86	84	87	94	BIC	33	75	89	86	87	90	**96**	96
KICc	89	**93**	84	80	88	80	58	41	KICc	**72**	82	90	77	76	78	62	45
GAIC	51	82	88	89	96	94	**94**	**95**	GAIC	3	31	50	65	76	76	93	**99**

Table 2. The counts indicate for 100 runs the number of times the number of sinusoids was correctly estimated by each criterion. The best result for each sample size N is represented with bold font.

In our settings, the maximum number of sinusoids is $K_{max} = 8$, and the maximum order of the AR process depends on the number of the available measurements: $M_{max} = \lfloor \ln^2 N \rfloor - 1$. The formula for M_{max} is derived from the condition used in [2] to ensure the consistency of the BIC criterion. Supplementarily, each pair (K, M) must verify the inequality $3K + M < N - 2$ to be a candidate for the model structure.

Examples 1-3 are taken from [2], where the estimation results are reported only for $N \geq 128$. Since our main interest is on small and moderate sample sizes, we evaluate the performances of the information theoretic criteria for $N \in \{25, \ldots, 100\}$ and various levels of the local SNR. In Examples 1-3, we consider $K = 2$ sinusoids whose parameters are $\xi_1 = (2^{1/2}, 1, 0)$ and $\xi_2 = (2^{-1/2}, 2, 0)$. The additive noise is generated as follows:
Example 1: $e_t = \varepsilon_t$ (white noise),
Example 2: $e_t = -0.81 e_{t-2} + \varepsilon_t$ (autoregressive noise),
Example 3: $e_t = \varepsilon_t + 1.6\varepsilon_{t-1} + 0.64\varepsilon_{t-2}$ (moving average noise),
where ε_t is a sequence of i.i.d Gaussian random variables with zero mean and variance σ^2, which is chosen such that the local SNR's take the desired values.
Example 4 is taken from [3] and modified such that the observations y^N are real-valued. The number of sinusoids is $K = 3$ and their parameters are $\xi_1 = (2, 0.10\pi, 0)$, $\xi_2 = (2, 0.80\pi, 0)$ and $\xi_3 = (2, 0.84\pi, 0)$. The noise is simulated by the autoregressive process $e_t = 0.85 e_{t-1} + \varepsilon_t$, where the significance of ε_t is the same as above.

We focus on the capabilities of the tested criteria to estimate correctly the number of sinusoids K. For the Examples 1-4, we count the number of correct estimates for 100 runs when the local SNR's and the sample size N take various values. The results are reported in Table 2.

Remark that SCp is the best among the SC formulae and its performances are closely followed by SCa. For both SCp and SCa, FIM of the sinusoidal components are decoupled [1], which is a serious computational advantage. From the results reported in [1], we can draw the conclusion that the shape of the noise spectrum has more influence on SCp than on SCa, and this explains the superiority of the SCp criterion. The performances of SCe are very modest because FIM used in SCe can be ill-conditioned for small and moderate sample size when the number of sinusoids is two or larger [1].

When the sample size N is smaller than 80, SCa is superior to BIC and GAIC. This is a straightforward consequence of the asymptotic approximations applied in the derivations of the BIC and GAIC criteria. KICc estimates for the number of sinusoids are remarkably correct when $N \leq 40$, but the number of correct estimations yield by KICc declines when N increases such that for $N \geq 80$ the reported results are very modest.

We extend our analysis by counting the Type I and Type II errors. Let $f_k = \omega_k/(2\pi)$ and similarly $\hat{f}_k = \hat{\omega}_k/(2\pi)$. Since K and $\hat{K}$ are not necessarily equal, we take $\mathcal{K} = \min(K, \hat{K})$. We select the indices $\{i_1, \ldots, i_{\mathcal{K}}\} \subseteq \{1, \ldots, K\}$ and $\{j_1, \ldots, j_{\mathcal{K}}\} \subseteq \{1, \ldots, \hat{K}\}$ such that $|f_{i_1} - \hat{f}_{j_1}|, \ldots, |f_{i_{\mathcal{K}}} - \hat{f}_{j_{\mathcal{K}}}|$ are the smallest entries of the set $\{|f_i - \hat{f}_j| \, : \, 1 \leq i \leq K, 1 \leq j \leq \hat{K}\}$. For each $k \in \{1, \ldots, \mathcal{K}\}$, $\hat{f}_{j_k}$ is deemed to be the estimate for f_{i_k}. As usual, a Type I error is counted in connection with the frequency f_k if none of the estimated frequencies are assigned to f_k, and a Type II error is counted whenever $\hat{K} > K$. We compute also the mean-squared errors (MSE) for the frequency estimates.

For brevity, we report in Table 3 the Type I and Type II errors together with the MSE only for the experiments conducted in Example 2 when SNR_2=3.00 dB. In our comparisons, we consider SCp and the asymptotic criteria BIC and GAIC.

We note that only GAIC has difficulties in recovering the first harmonic when $N < 50$, and recovering the second harmonic whose local SNR is smaller posses problems to all the criteria. Remark for SCp that the number of Type I errors connected with f_2 decreases very fast with the increase of the sample size. For GAIC, the number of Type II errors is always small, but many Type I errors occur even for $N = 60$. This is a clear sign that, for small N, GAIC underestimates the number of sinusoids. The computed MSE is almost the same for all the investigated criteria and this is natural because the evaluation of SCp, BIC and GAIC is based on the estimates provided by the RELAX algorithm.

Conclusion The use of approximations for SC that are sharper than BIC improves the estimation results for small and moderate sample size. The results encourage us to extend the application of SC to other classes of models like 1-D damped harmonics in colored noise, or 2-D sinusoids in white noise.

Appendix:On the derivation of SC formula (3)

To check the conditions for the applicability of the SC formula of Qian and Künsch in the particular case of the model given by the equations (1) and (2), we resort to a closed-form expression of $\mathbf{J}_N(\theta)$ that was derived for large sample size [4]. When N is large, under both H_{dp} and Hrp, $\mathbf{J}_N(\theta)$ is block-diagonal such that the block $\mathbf{J}_N(\xi_k)$ corresponds to the parameters of the k-th sine-wave and the block $\mathbf{J}_N(\mathfrak{a})$ corresponds to the parameters of the AR noise [1]. More precisely, we have $\mathbf{J}_N(\xi_k) = \mathbf{Q}_N \mathbf{G}(\xi_k, \mathfrak{a}) \mathbf{Q}_N$, where $\mathbf{Q}_N = \text{diag}(N^{1/2}\, N^{3/2}\, N^{1/2})$ and $\mathbf{G}(\xi_k, \mathfrak{a}) = \text{SNR}_k \begin{bmatrix} 1/\alpha_k^2 & 0 & 0 \\ 0 & 1/3 & 1/2 \\ 0 & 1/2 & 1 \end{bmatrix}$. Here SNR_k denotes the local SNR for the k-th sinusoidal component. The entries of $\mathbf{J}_N(\mathfrak{a})$ are not influenced by the parameters ξ, hence $\mathbf{J}_N(\mathfrak{a})$ is the same as in the pure AR case. Based on results from [12] we can write $\mathbf{J}_N(\mathfrak{a}) = \frac{N}{\tau}\mathbf{R}(\mathfrak{a})$, where $\mathbf{R}(\mathfrak{a})$ is the $M \times M$ covariance matrix of the AR process defined in (2).

After these preliminaries, we outline the conditions as they are given in [8]: (C1) $\mathbf{J}_N(\theta)$ is positive definite; (C2) The minimum eigenvalue of $\mathbf{J}_N(\theta)$ is of order $O(N)$ as $N \to \infty$;
(C3) $|\mathbf{J}_N(\theta_1)|^{-1} ||\mathbf{J}_N(\theta_1)| - |\mathbf{J}_N(\theta_2)|| \le c||\theta_1 - \theta_2||$, $\forall \theta_1, \theta_2 \in \Theta$, where c is a finite constant; (C4) $\ln|\mathbf{J}_N(\theta)| = o(N)$.

It is easy to check that all the eigenvalues of $\mathbf{J}_N(\xi_k)$ are strict positive. The covariance matrix $\mathbf{R}(\mathfrak{a})$ is positive definite for any M, therefore $\mathbf{J}_N(\mathfrak{a})$ is also positive definite, and the condition C1 is verified. Two of the eigenvalues of $\mathbf{J}_N(\xi_k)$ are $O(N)$ and the third one is $O(N^3)$. As each eigenvalue of $\mathbf{J}_N(\mathfrak{a})$ is $O(N)$, we conclude that C2 is satisfied. A longer discussion is necessary for C3, but we only remark here that $|\mathbf{J}_N(\theta)| = N^{5K+M} \frac{|\mathbf{R}(\mathfrak{a})|}{\tau^M} \prod_{k=1}^{K} |\mathbf{G}(\xi_k, \mathfrak{a})|\ \forall\theta$, thus the left-hand-side term in the inequality C3 is finite and does not depend on N. Using the expression above for $|\mathbf{J}_N(\theta)|$, we readily obtain $\lim_{N\to\infty} \frac{\ln|\mathbf{J}_N(\theta)|}{N} = 0$, thus C4 is verified.

We can apply next the SC formula from [8]. For simplicity we ignore the terms that do not depend on N, and the formula becomes:

$$-\log f(y^N; \hat{\theta}, \hat{\tau}) + \log|\tilde{\mathbf{J}}_N(\hat{\theta}, y^N)|^{1/2}$$
$$+\sum_{i=1}^{3K+M} \log(|\hat{\theta}_i| + N^{-1/4}) + \sum_{i=1}^{3K+M} r^*(N^{1/4}|\hat{\theta}_i| + 1) + O(N^{-1/4}),$$

where $\log(\cdot)$ is the logarithm base 2, $\hat{\theta}$ denotes the ML estimates, and $\tilde{\mathbf{J}}_N(\hat{\theta}, y^N) = -\frac{\partial^2 \ln f(y^N; \theta, \tau)}{\partial\theta\partial\theta^\top}\big|_{\theta=\hat{\theta}}$ is the observed FIM.

Frequency f_1 : Err. I								
N	30	35	40	45	50	60	80	100
SCp	2	5	5	0	0	0	0	0
BIC	3	12	15	5	0	0	0	0
GAIC	28	31	42	26	11	8	0	0
Frequency f_1 : MSE								
N	30	35	40	45	50	60	80	100
SCp	-59.49	-59.08	-62.82	-63.87	-64.23	-66.57	-75.42	-80.15
BIC	-59.78	-59.15	-62.03	-64.57	-64.02	-66.99	-75.42	-80.15
GAIC	-59.07	-59.96	-62.60	-64.88	-63.58	-67.88	-75.42	-80.15
Frequency f_2 : Err. I								
N	30	35	40	45	50	60	80	100
SCp	29	23	12	4	1	0	0	0
BIC	28	29	28	23	4	6	0	0
GAIC	86	82	79	75	48	32	2	0
Frequency f_2 : MSE								
N	30	35	40	45	50	60	80	100
SCp	-50.29	-52.24	-56.82	-55.97	-57.89	-60.81	-63.64	-67.16
BIC	-50.67	-39.02	-57.57	-57.70	-57.82	-41.61	-63.56	-67.16
GAIC	-51.31	-50.60	-58.75	-57.40	-57.28	-40.22	-63.96	-67.16
Err. II								
N	30	35	40	45	50	60	80	100
SCp	14	23	14	13	5	11	5	4
BIC	41	39	28	18	12	10	3	2
GAIC	3	11	6	4	1	2	2	2

Table 3. Type I and Type II errors for Example 2 when SNR_2=3.00 dB. MSE is computed for the estimates of the frequencies and it is expressed in dB. The results are reported for 100 runs.

For any $x > 0$, $r^*(x) = \log(\log x) + \log(\log(\log x)) + \cdots$, where the sum continues as long as the iterated logarithms are strict positive.

The approximative formula (3) is obtained from the expression above after replacing $\tilde{\mathbf{J}}_N(\hat{\theta}, y^N)$ with $\mathbf{J}_N(\hat{\theta})$, discarding an $O((3K + M)\log\log N)$ term, and changing $\log(\cdot)$ to $\ln(\cdot)$.

It is recommended in [8] to consider in the SC expression also the term given by the number of parameters divided by two and multiplied by $\log\rho$, where ρ is the largest eigenvalue of $\mathbf{J}_N(\hat{\theta})^{-1/2}\tilde{\mathbf{J}}_N(\hat{\theta}, y^N)\mathbf{J}_N(\hat{\theta})^{-1/2}$. It can be readily verified that under mild conditions ρ does not depend on N, hence we ignore the $\log\rho$ term in (3).

5. REFERENCES

[1] M. Ghogho and A. Swami, "Fast computation of the exact FIM for deterministic signals in colored noise," *IEEE Trans. Signal. Proces.*, vol. 47, no. 1, pp. 52–61, Jan. 1999.

[2] L. Kavalieris and E.J. Hannan, "Determining the number of terms in a trigonometric regression," *Journal of time series analysis*, vol. 15, no. 6, pp. 613–625, 1994.

[3] J. Li and P. Stoica, "Efficient mixed-spectrum estimation with applications to target feature extraction," *IEEE Trans. Signal. Proces.*, vol. 44, no. 2, pp. 281–295, Feb. 1996.

[4] P. Stoica and A. Nehorai, "Statistical analysis of two nonlinear least-squares estimators of sine-wave parameters in the colored-noise case," *Circuits, Systems, and Signal Processing*, vol. 8, no. 1, pp. 3–15, 1989.

[5] P. Stoica and Y. Selen, "A review of information criterion rules," *IEEE Signal. Proces. Mag.*, vol. 21, no. 4, pp. 36–47, Jul. 2004.

[6] J. Rissanen, "Fisher information and stochastic complexity," *IEEE Trans. Inf. Theory*, vol. 42, no. 1, pp. 40–47, Jan. 1996.

[7] J. Rissanen, "MDL denoising," *IEEE Trans. Inf. Theory*, vol. 46, no. 7, pp. 2537–2543, Nov. 2000.

[8] G. Qian and H.R. Kunsch, "Some notes on Rissanen's stochastic complexity," *IEEE Trans. Inf. Theory*, vol. 44, no. 2, pp. 782–786, Mar. 1998.

[9] C.D. Giurcaneanu and J. Rissanen, "Estimation of AR and ARMA models by stochastic complexity," in *Time series and related topics.*, Hwai-Chung Ho, Ching-Kang Ing, and Tze Leung Lai, Eds., vol. 52, pp. 48–59. Institute of Mathematical Statistics Lecture Notes-Monograph Series, 2006.

[10] P. M. Djuric, "A model selection rule for sinusoids in white Gaussian noise," *IEEE Trans. Signal. Proces.*, vol. 44, no. 7, pp. 1744–1751, Jul. 1996.

[11] B. G. Quinn, "Estimating the number of terms in a sinusoidal regression," *Journal of time series analysis*, vol. 10, no. 1, pp. 70–75, 1989.

[12] B. Friedlander and B. Porat, "The exact Cramer-Rao bound for Gaussian autoregressive processes," *IEEE Tr. on Aerospace and Electronic Systems*, vol. AES-25, pp. 3–8, 1989.

[13] S.P. Bruzzone and M. Kaveh, "Information tradeoffs in using the sample autocorrelation function in ARMA parameter estimation," *IEEE Trans. on Acoustics, Speech and Signal Processing*, vol. ASSP-32, no. 4, pp. 701–715, Aug. 1984.

[14] A.-K. Seghouane and M. Bekara, "A small sample model selection criterion based on Kullback's symmetric divergence," *IEEE Trans. Signal. Proces.*, vol. 52, pp. 3314–3323, 2004.

[15] M. Bekara, L. Knockaert, A.-K. Seghouane, and G. Fleury, "A model selection approach to signal denoising using Kullback's symmetric divergence," *Signal Processing*, vol. 86, pp. 1400–1409, 2006.