Using Information Theoretic Techniques for Sinusoidal Signal Resolution

S. Valaee[1], *B. Champagne*[2], *P. Kabal*[3]

[1]Dept. of Elect. Eng., Tarbiat Modares University, Tehran, Iran
[2]INRS-Telecommunications, Verdun, Quebec, Canada
[3]Dept. of Elect. Eng., McGill University, Montreal, Canada

Abstract

The objective is to develop information theoretic criteria for detection of sinusoidal signals. The minimum description length (MDL) and the predictive stochastic complexity (PSC) have been formulated for harmonic resolution. MDL and PSC are the codelength for data and model. The proposed techniques are based on decomposing the observation vector into its components in the signal and noise subspaces. Each component is encoded separately and the results are added to form the total codelength. The codelength is minimized over different models to select the best model.

1 Introduction

Sinusoidal signal detection is applied in various fields ranging from telecommunications to array processing and spectrum estimation. Various tecchniques have been proposed in the literature based on the low resolution as well as the high resolution approaches; see [1] and [2]. In some techniques, it is frequently assumed that the number of signals is known. This is an unrealistic assumption which might not hold in practice.

Here, we propose two techniques that can be used for signal enumeration. The techniques are based on the information theoretic approach. We formulate the problem based on the minimum description lenght (MDL) [3] and the predictive stochastic complexity (PSC) [4] principles. Both techniques are used to estimate the model order by minimizing the Kullback-Leibler distance between the true model and the estimated one.

Direct application of MDL and PSC to sinusoidal resolution generates erroneous results — the number of signals is always detected as 1. This is due to the temporal coherency of sinusoids. In the present work, we introduce an alternative approach, such as the one presented in [5] and [6]. The proposed technique is based on decomposing the observation vectors into their orthogonal components in the signal and noise subspaces. Using the MDL or PSC principle, these components are encoded separately and the results are added to obtain the total codelength. This procedure is performed for all possible models and the minimum codelength is selected to give the best model. Simulation study shows that both techniques can detect the number of signals. PSC has a better performancce in nonstationary environments.

2 Problem Formulation

Assume that the observation is a time series modeled at time t as

$$x(t) = \sum_{k=1}^{K} \alpha_k \cos(\omega_k t + \phi_k) + n(t) \qquad (1)$$

where the parameters $\theta^k = (\alpha_k, \omega_k, \phi_k), k = 1, \ldots, K$, and their number K are unknown; $n(t)$ is a Gaussian white noise with an unknown variance σ^2. All unknowns can be arranged in a parameter vector

$$\Psi = (\alpha_1, \omega_1, \phi_1, \ldots, \alpha_K, \omega_K, \phi_K, \sigma^2). \qquad (2)$$

The observed data is sampled with the sampling rate $\omega_s > 2\max_k\{\omega_k\}$ and arranged in a matrix form with each column representing an $M \times 1$ snapshot vector

$$\mathbf{x}(t) = \sum_{k=1}^{K} \mathbf{a}(\omega_k)\mathbf{s}(t, \alpha_k, \omega_k, \phi_k) + \mathbf{n}(t) \qquad (3)$$

where

$$\mathbf{a}(\omega_k)=\begin{bmatrix} 1 & 0 \\ \cos(\omega_k D) & \sin(\omega_k D) \\ \vdots & \vdots \\ \cos(\omega_k(M-1)D) & \sin(\omega_k(M-1)D) \end{bmatrix} \quad (4)$$

with $D = \frac{2\pi}{\omega_s}$ being the sampling interval, and

$$\mathbf{s}(t,\alpha_k,\omega_k,\phi_k) = \begin{bmatrix} \alpha_k \cos(\omega_k t + \phi_k) \\ -\alpha_k \sin(\omega_k t + \phi_k) \end{bmatrix}. \quad (5)$$

The matrix $\mathbf{a}(\omega_k)$ is time-invariant — it is only a function of the frequency ω_k. Arrangement of all $\mathbf{a}(\omega_k), k = 1,\ldots,K$ in a matrix gives

$$\mathbf{A}(\Omega) = [\mathbf{a}(\omega_1),\ldots,\mathbf{a}(\omega_K)] \quad (6)$$

where $\Omega = (\omega_1,\ldots,\omega_K)$ is the vector of all frequencies of the sinusoids. The *signal subspace* is defined as the span of $\mathbf{A}(\Omega)$. The *noise subspace* is the orthogonal complement of the signal subspace.

Let $\mathbf{X}(T) = [\mathbf{x}(t)], t = 1,\ldots,T$, be the $M \times T$ observation matrix — the matrix of snapshot vectors collected in the window $(1,\ldots,T)$. Using the observation matrix $\mathbf{X}(T)$, we present information theoretic methods to estimate the number of signals K and their frequencies $\omega_k, k = 1,\ldots,K$.

3 Information Theoretic Criteria

We use the *minimum description length* (MDL) [3] and the *predictive stochastic complexity* (PSC) [4] techniques. PSC and MDL are the codelengths used to represent data. Both principles are based on minimizing the Kullback-Leibler distance between the true model and the estimated one.

The MDL criterion for a model of order k at time instant T is

$$\mathrm{MDL}(T,k) = -\log f\left(\mathbf{X}(T)|\hat{\mathbf{\Psi}}_T^k\right) + \frac{k}{2}\log T \quad (7)$$

where $f(\mathbf{X}|\mathbf{\Psi})$ is the conditional probability density function, and $\hat{\mathbf{\Psi}}_T^k$ is the *maximum likelihood* (ML) estimate of the parameter vector $\mathbf{\Psi}^k$ using the observations up to time T. In MDL, data and model are encoded separately and the results are added to obtain the total codelength. The model order at time T is determined from

$$\hat{K} = \min_k \mathrm{MDL}(T,k) \quad (8)$$

where the minimization is performed over all possible models.

PSC is the codelength for a minimal description of data; at time T and for a model order k, it amounts to

$$\mathrm{PSC}(T,k) = -\sum_{t=1}^{T} \log f\left(\mathbf{x}(t)|\hat{\mathbf{\Psi}}_{t-1}^k\right) \quad (9)$$

where $\hat{\mathbf{\Psi}}_{t-1}^k$ is the ML estimate of the parameter vector $\mathbf{\Psi}^k$ using the observations up to time $(t-1)$. The estimated model order at time T is given by

$$\hat{K} = \min_k \mathrm{PSC}(T,k) \quad (10)$$

with the minimization performed over all possible models.

4 Harmonic Resolution

In a straightforward approach, the conditional probability density of $\mathbf{X}(T)$ is determined and used in (7) and (9). This approach to detection of sinusoids produces erroneous results — in fact the model order is always estimated as 1. This is due to the temporal coherency of the signals.

In this paper, we take an alternative approach similar to the one presented in [5] [6]. We propose decomposing the observation vectors into their components in the signal and noise subspaces and encoding them separately. Since the components of the observation vectors in the signal and noise subspaces are independent, the total codelength will be the sum of the codelengths of the two components.

Let us represent by $\mathbf{P}_s(\Omega)$ and $\mathbf{P}_n(\Omega)$ the projection matrices onto the signal and noise subspaces, respectively. The signal subspace is the column span of $\mathbf{A}(\Omega)$, hence the projection matrix onto the signal subspace is given by

$$\mathbf{P}_s(\Omega) = \mathbf{A}(\Omega)\left(\mathbf{A}^H(\Omega)\mathbf{A}(\Omega)\right)^{-1}\mathbf{A}^H(\Omega). \quad (11)$$

The projection matrix onto the noise subspace is then

$$\mathbf{P}_n(\Omega) = \mathbf{I} - \mathbf{P}_s(\Omega), \quad (12)$$

where $\mathbf{I}$ is the $M \times M$ unity matrix. The observation vector $\mathbf{x}(t)$ can be decomposed as

$$\mathbf{x}(t) = \mathbf{P}_s(\Omega)\mathbf{x}(t) + \mathbf{P}_n(\Omega)\mathbf{x}(t). \quad (13)$$

The $M \times 1$ vector $\mathbf{P}_s(\Omega)\mathbf{x}(t)$ is in the $2K$-dimensional signal subspace. Similarly, $\mathbf{P}_n(\Omega)\mathbf{x}(t)$

is in the $(M-2K)$-dimensional noise subspace. We represent these vectors by

$$\mathbf{x}_s(t) = \mathbf{P}_s(\Omega)\mathbf{x}(t), \quad (14)$$
$$\mathbf{x}_n(t) = \mathbf{P}_n(\Omega)\mathbf{x}(t). \quad (15)$$

Note that $\mathbf{x}_s(t)$ and $\mathbf{x}_n(t)$ are orthogonal. Hence

$$\mathbf{R}_x = \mathbf{R}_s(\Omega) + \mathbf{R}_n(\Omega) \quad (16)$$

where $\mathbf{R}_x$ is the correlation matrix of the observation vector $\mathbf{x}(t)$, and

$$\mathbf{R}_s(\Omega) = \mathbf{P}_s(\Omega)\mathbf{R}_x\mathbf{P}_s(\Omega), \quad (17)$$
$$\mathbf{R}_n(\Omega) = \mathbf{P}_n(\Omega)\mathbf{R}_x\mathbf{P}_n(\Omega) \quad (18)$$

are the correlation matrices of the signal and noise components, respectively.

The noise vector $n(t)$ is a white Gaussain process. The probabilty density function of $\mathbf{x}(t)$ is then

$$f(\mathbf{x}(t)|\mathbf{R}_x) = |\pi\mathbf{R}_x|^{-1}\exp\{-\mathbf{x}^H(t)\mathbf{R}_x^{-1}\mathbf{x}(t)\}. \quad (19)$$

We use a stochastic modeling for the signals — an example would be when α or ϕ are stochastic processes. The probability density function for $\mathbf{X}(T)$ is

$$f(\mathbf{X}(T)|\mathbf{R}_x) = |\pi\mathbf{R}_x|^{-T}\exp\{-T\mathrm{tr}(\mathbf{R}_x^{-1}\bar{\mathbf{R}}_x)\} \quad (20)$$

where tr(.) is the trace operator and

$$\bar{\mathbf{R}}_x = \frac{1}{T}\sum_{t=1}^{T}\mathbf{x}(t)\mathbf{x}^H(t) \quad (21)$$

is the sample correlation matrix of the observation vector. The log-likelihood function for $\mathbf{X}(T)$ is

$$-\log f(\mathbf{X}(T)|\mathbf{R}_x) = T\log|\pi\mathbf{R}_x| + T\mathrm{tr}(\mathbf{R}_x^{-1}\bar{\mathbf{R}}_x). \quad (22)$$

We develope the MDL and PSC information theoretic criteria for signal resolution. The following lemmas will be used later.

Lemma 1 *Let* $\mathbf{A}$, $\mathbf{B}$ *be* $n \times n$ *Hermitian matrices orthogonal to each other such that* $\mathbf{A}^H\mathbf{B} = \mathbf{B}^H\mathbf{A} = 0$. *If the matrix* $\mathbf{C}$ *is given by*

$$\mathbf{C} = \mathbf{A} + \mathbf{B} \quad (23)$$

where $\mathbf{C}$ *is full rank, then*

$$|\mathbf{C}| = \zeta(\mathbf{A})\,\zeta(\mathbf{B}) \quad (24)$$

where $|.|$ *is the determinant, and* $\zeta(.)$ *represents the multiplication of the nonzero eigenvalues.*

Proof: See Appendix A.

Lemma 2 *Let* $\lambda_1 > \ldots > \lambda_r$ *with multiplicities* $m_1, \ldots, m_r$ *be the eigenvalues of correlation matrix* $\mathbf{R}_x$, *and* v_k *be the eigenvectors. Corresponding values for the sample correlation matrix are defined by* $\bar{\lambda}_k$ *and* $\bar{v}_k$. *Then the ML estimate of* λ_k *is*

$$\hat{\lambda}_k = \frac{1}{m_k}\sum_{i\in L_k}\bar{\lambda}_i \quad (25)$$

where L_k *is the set of integers* $\{\sum_{j=1}^{k-1}m_j + 1, \ldots, \sum_{j=1}^{k-1}m_j + m_k\}$, *and the ML estimate of* v_k *is*

$$\hat{v}_k = \bar{v}_k. \quad (26)$$

Proof: See [7].

4.1 The MDL criterion

The log-likelihood function using the ML estimate of the correlation matrix is given by

$$-\log f(\mathbf{X}(T)|\hat{\mathbf{R}}_x) = TM\log\pi + T\log|\hat{\mathbf{R}}_x| + T\mathrm{tr}(\hat{\mathbf{R}}_x^{-1}\bar{\mathbf{R}}_x) \quad (27)$$

where $\hat{\mathbf{R}}_x$ is the ML estimate of the correlation matrix $\mathbf{R}_x$. Let us represent by $\lambda_i(R)$ and $\mathbf{v}_i(R)$ the eigenvalues and the corresponding eigenvectors of the correlation matrix R, with the eigenvalues arranged in non-increasing order. Using Lemma 2, we have

$$\mathbf{v}_i(\hat{\mathbf{R}}_x) = \mathbf{v}_i(\bar{\mathbf{R}}_x), \quad i = 1, \ldots, M \quad (28)$$
$$\lambda_i(\hat{\mathbf{R}}_x) = \lambda_i(\bar{\mathbf{R}}_x), \quad i = 1, \ldots, 2K \quad (29)$$
$$\lambda_i(\hat{\mathbf{R}}_x) = \frac{1}{M-2K}\sum_{k=2K+1}^{M}\lambda_k(\bar{\mathbf{R}}_x), \quad (30)$$
$$i = 2K+1, \ldots, M.$$

Using these results,

$$\mathrm{tr}(\hat{\mathbf{R}}_x^{-1}\bar{\mathbf{R}}_x) = M. \quad (31)$$

Substituting this in (27) gives

$$-\log f(\mathbf{X}(T)|\hat{\mathbf{R}}_s + \hat{\mathbf{R}}_n) = TM\log\pi + T\log|\hat{\mathbf{R}}_s + \hat{\mathbf{R}}_n| + TM. \quad (32)$$

We know find the number of unknowns that can be freely chosen. $\hat{\mathbf{R}}_s$ is a complex Hermitiam matrix with rank $2K$. Thus, the number of free parameters in determining $\hat{\mathbf{R}}_s$ is $4K^2$. The parameter vector Ω is determined by K parameters, and

$\hat{\sigma}^2$ is an unknown scalar. Thus the MDL criterion, excluding the terms independent of the model order K, is given by

$$\begin{aligned} \mathrm{MDL}(T,K) &= T\log|\hat{\mathbf{R}}_s+\hat{\mathbf{R}}_n| \\ &\quad +\frac{4K^2+K}{2}\log T. \end{aligned} \quad (33)$$

Using Lemma 1, we have

$$\begin{aligned} \mathrm{MDL}(T,K) &= T\log\left(\zeta(\hat{\mathbf{R}}_s)\zeta(\hat{\mathbf{R}}_n)\right) \\ &\quad +\frac{4K^2+K}{2}\log T. \end{aligned} \quad (34)$$

Let us define

$$\bar{\mathbf{R}}_s = \mathbf{P}_s(\Omega)\bar{\mathbf{R}}_x\mathbf{P}_s(\Omega), \quad (35)$$

$$\bar{\mathbf{R}}_n = \mathbf{P}_n(\Omega)\bar{\mathbf{R}}_x\mathbf{P}_n(\Omega). \quad (36)$$

Using these definitions,

$$\zeta(\hat{\mathbf{R}}_s) = \prod_{i=1}^{2K}\lambda_i(\bar{\mathbf{R}}_s), \quad (37)$$

$$\zeta(\hat{\mathbf{R}}_n) = \left(\frac{1}{M-2K}\sum_{i=1}^{M-2K}\lambda_i(\bar{\mathbf{R}}_n)\right)^{(M-2K)} \quad (38)$$

So far we have assumed that K and Ω are known, whereas in practice they are to be estimated. Define

$$\Omega^k = (\omega_1,\ldots,\omega_k) \quad (39)$$

where $\omega_1,\ldots,\omega_k$ are unknown frequencies of the sinusoids. We find MDL criterion for all $k = 0,\ldots,M/2-1$ and choose the minimum to detect the number of signals. The MDL criterion for the model k is then given by

$$\begin{aligned} \mathrm{MDL}(T,k) &= T\log\left(\zeta(\hat{\mathbf{R}}_s^k)\zeta(\hat{\mathbf{R}}_n^k)\right) \\ &\quad +\frac{4k^2+k}{2}\log T \end{aligned} \quad (40)$$

where

$$\zeta(\hat{\mathbf{R}}_s^k) = \prod_{i=1}^{2k}\lambda_i(\bar{\mathbf{R}}_s^k), \quad (41)$$

$$\zeta(\hat{\mathbf{R}}_n^k) = \left(\frac{1}{M-2k}\sum_{i=1}^{M-2k}\lambda_i(\bar{\mathbf{R}}_n^k)\right)^{(M-2k)} \quad (42)$$

and

$$\bar{\mathbf{R}}_s^k = \mathbf{P}_s(\Omega^k)\bar{\mathbf{R}}_x\mathbf{P}_s(\Omega^k), \quad (43)$$

$$\bar{\mathbf{R}}_n^k = \mathbf{P}_n(\Omega^k)\bar{\mathbf{R}}_x\mathbf{P}_n(\Omega^k). \quad (44)$$

As seen the computation of the MDL criterion depends on the parameter vector Ω^k. In the original MDL approach, an ML estimate of Ω^k is required.

4.2 The PSC criterion

The PSC criterion is computed for all t inside the window $[0,T]$. The final value is minimized at the end of the window (although the minimization can be performed at each step).

Let us define the sample correlation matrix at time instant t by

$$\bar{\mathbf{R}}_{x_t} = \frac{1}{t}\sum_{i=1}^{t}\mathbf{x}(i)\mathbf{x}^H(i). \quad (45)$$

The projection of this matrix onto the signal and noise subspaces for model k are defined as

$$\bar{\mathbf{R}}_{s_t}^k = \mathbf{P}_s(\Omega_t^k)\bar{\mathbf{R}}_{x_t}\mathbf{P}_s(\Omega_t^k), \quad (46)$$

$$\bar{\mathbf{R}}_{n_t}^k = \mathbf{P}_n(\Omega_t^k)\bar{\mathbf{R}}_{x_t}\mathbf{P}_n(\Omega_t^k), \quad (47)$$

where Ω_t^k is the estimate of parameter vector for model k using the observations upto time t.

The ML estimate of the correlation matrix for the k-th model and the $(t-1)$-th snapshot is

$$\hat{\mathbf{R}}_{x_{t-1}}^k = \hat{\mathbf{R}}_{s_{t-1}}^k + \hat{\mathbf{R}}_{n_{t-1}}^k \quad (48)$$

where $\hat{\mathbf{R}}_{s_{t-1}}^k$ and $\hat{\mathbf{R}}_{n_{t-1}}^k$ are the ML estimate of the projection of correlation matrix onto the signal and noise subspaces. Using Lemma 2,

$$\hat{\mathbf{R}}_{s_{t-1}}^k = \bar{\mathbf{R}}_{s_{t-1}}^k. \quad (49)$$

Similary, it is possible to show that $\hat{\mathbf{R}}_{n_{t-1}}^k$ has the same eigenvectors as $\bar{\mathbf{R}}_{n_{t-1}}^k$ and a single eigenvalue with multiplicity $(M-2k)$ which is found from

$$\hat{\sigma}^2(\Omega_{t-1}^k) = \frac{1}{M-2k}\mathrm{tr}\bar{\mathbf{R}}_{n_{t-1}}^k. \quad (50)$$

Note that $\hat{\mathbf{R}}_{n_{t-1}}^k$ can be obtained by

$$\hat{\mathbf{R}}_{n_{t-1}}^k = \mathbf{T}_{t-1}^k\bar{\mathbf{R}}_{n_{t-1}}^k \quad (51)$$

where $\mathbf{T}_{t-1}^k$ is a matrix defined as

$$\mathbf{T}_{t-1}^k = \bar{\mathbf{V}}_{n,M-2k}\,\mathrm{diag}\left[\frac{\hat{\sigma}^2(\Omega_{t-1}^k)}{\lambda_j(\bar{\mathbf{R}}_{n_{t-1}})}\right]\bar{\mathbf{V}}_{n,M-2k}^H \quad (52)$$

with $\lambda_j(\bar{\mathbf{R}}_{n_{t-1}})$, $j = 1,\ldots,M-2k$, being the nonzero eigenvalues of $\bar{\mathbf{R}}_{n_{t-1}}$, and $\bar{\mathbf{V}}_{n,M-2k}$, the $M\times(M-2k)$ matrix of corresponding eigenvectors; the diag[.] is a representation for a diagonal matrix formed by the elements in the brackets.

Using Lemma 1, the ML estimate of the determinant of the correlation matrix is obtained by the multiplication of the nonzero eigenvalues of its projected components

$$|\hat{\mathbf{R}}_{x_{t-1}}^k| = \zeta(\hat{\mathbf{R}}_{s_{t-1}}^k)\,\zeta(\hat{\mathbf{R}}_{n_{t-1}}^k) \quad (53)$$

where from (49) and (50),

$$\zeta(\hat{\mathbf{R}}^k_{s_{t-1}}) = \zeta(\bar{\mathbf{R}}^k_{s_{t-1}}), \quad (54)$$

$$\zeta(\hat{\mathbf{R}}^k_{n_{t-1}}) = \left(\hat{\sigma}^2(\Omega^k_{t-1})\right)^{M-2k}. \quad (55)$$

Using these results, the PSC criterion is

$$\begin{aligned}\mathrm{PSC}_k(N) = \sum_{t=1}^{T} \Big[& \log\zeta(\bar{\mathbf{R}}^k_{s_{t-1}}) \\ & +(M-2k)\log(\frac{1}{M-2k}\mathrm{tr}\bar{\mathbf{R}}^k_{n_{t-1}}) \\ & +\mathbf{x}_t^H(\bar{\mathbf{R}}^k_{s_{t-1}} + \mathbf{T}^k_{t-1}\bar{\mathbf{R}}^k_{n_{t-1}})^{-1}\mathbf{x}_t\Big]. \end{aligned} \quad (56)$$

It is seen that the computation of PSC depends on the parameter vector Ω^k_{t-1}. In the original version of the PSC algorithm the ML estimate of the parameter vector is used.

5 Simulation Results

We include here the results for the simulation study. To avoid the computational complexity of the ML estimator, we choose to use a root MUSIC technique to estimate Ω^k.

Example 1: We study a scenario with two sinusoids with the parameters: $\{\alpha_1 = 2, \omega_1 = 110, \phi_1 = \frac{\pi}{3}\}$, and $\{\alpha_2 = 1, \omega_2 = 160, \phi_2 = -\frac{\pi}{4}\}$. The sampling interval is 1 ms. The data were collected over 0.8 second and decomposed into 50 non-overlapping snapshots of length 16 samples each. The case was simulated for 100 independent trials. Table 1 compares the PSC and the MDL techniques based on the number of times that each method resolves the two signals as the noise power varies from −2 dB to 12 dB. The MDL and PSC algorithms have close performance.

Example 2: To study the two methods in a time varying environment, we simulate a case in which the phase of the first signal suddenly changes to $\frac{2\pi}{3}$ at $t = 570$ ms. The signals are such as in Example 1. The observation time is 1 second. The size of snapshot vector is 10. If the MDL criterion is computed and minimized at the end of observation, the number of signals will be dectected as 4. PSC is computed at each snapshot. The estimated model order as a function of the snapshots is depicted in Fig. 1. As seen PSC still detects signals. Fig. 2 illustrates the difference between the PSC terms. At $t = 570$ we notice an abrupt change in the PSC criterion. This sudden jump indicates that the statistics of the model has been altered. Thus, PSC can be used for change-point detection. MDL does not see this change — it is only calculated at the end of observation window.

Example 3: In this example we study a case in which the frequency of the second source is time-varying with a rate of 4 Hz per second. MDL and PSC are computed at each time instant. Note that usually MDL is not used as simulated here — we use it so as to compare the techniques based on their behavior to source drift. The results of model selection have been reported in Fig. 3. PSC breaks down much later than MDL. This is due to adaptive nature of PSC. In fact, at each time t, PSC adds a new term to the PSC criterion computed at the previous time instant for each model. This might compensate for the drift in the frequency. On the contrary, MDL uses all data upto time t and assumes that the characteristics of the sources are stationary.

These two examples show that the PSC algorithm might be more appropriate for a nonstationary environment.

6 Conclusion

This paper presented two information theoretic techniques for sinusoidal signal detection. The technqiues used the minimum description length (MDL) and the predictive stochastic complexity (PSC) principles. MDL and PSC both use a codelength of data for signal enumeration.

We used a subspace decomposition approach. Data were decomposed into their components in the signal and noise subspaces. Since the signal and noise subspaces were orthogonal, the total codelength of data was the addition of the codelength of their components. The codelengths were computed for each model and minimized over all models. Simulation study showed that for stationary environment the two techniques perform closely. For nonstationary environments PSC outperformed MDL.

A Proof of Lemma 1

Let Λ_a and Λ_b be the diagonal matrices of nonzero eigenvalues of $\mathbf{A}$ and $\mathbf{B}$, and $\mathbf{V}_a$ and $\mathbf{V}_b$ be the matrices of corresponding eigenvectors. Then $\mathbf{C}$ can be written as

$$\begin{aligned}\mathbf{C} &= \mathbf{V}_a\Lambda_a\mathbf{V}_a^H + \mathbf{V}_b\Lambda_b\mathbf{V}_b^H \\ &= \mathbf{V}\begin{bmatrix}\Lambda_a & 0 \\ 0 & \Lambda_b\end{bmatrix}\mathbf{V}^H \end{aligned} \quad (57)$$

Noise Power (dB)	Selected Model					
	PSC			MDL		
	1	2	3	1	2	3
−2	0	100	0	0	100	0
0	0	100	0	0	100	0
2	0	97	3	0	100	0
4	8	84	8	13	87	0
6	61	35	4	73	27	0
8	86	13	1	99	1	0
10	87	12	1	99	1	0
12	91	8	1	100	0	0

Table 1: The resolution of the two methods PSC and MDL.

where $\mathbf{V} = [\mathbf{V}_a \mathbf{V}_b]$. Since $\mathbf{C}$ is a full rank Hermitian matrix it is unitarily diagonalizable and the orthonormal matrix $\mathbf{V}$ is its eigenvector matrix . Thus the determinant of $\mathbf{C}$ is equal to

$$|\mathbf{C}| = |\Lambda_a|\,|\Lambda_b|. \tag{58}$$

Note that $|\Lambda_a|$ and $|\Lambda_b|$ are in fact the multiplication of the nonzero eigenvalues of $\mathbf{A}$ and $\mathbf{B}$.

References

[1] S. M. Kay and S. L. Marple. "spectrum analysis – a moderm perspective". *Proc. IEEE*, 69:1380–1419, Nov. 1981.

[2] S. B. Kesler. *Modern Spectrum Analysis, II.* IEEE Press, 1986.

[3] J. Rissanen. "modeling by shortest data description". *Automatica*, 14:465–471, 1979.

[4] J. Rissanen. "stochastic complexity and modeling". *Annals of Statistics*, 14:1080–1100, 1986.

[5] M. Wax. "detection of coherent and noncoherent signals via the stochastic signal model". In *IEEE Int. Conf. Acoust., Speech, Signal Processing*, pages 3541–3544, May 1991.

[6] S. Valaee and P. Kabal. "detection of the number of signals using predictive stochastic complexity". In *IEEE Int. Conf. Acoust., Speech, Signal Processing*, pages V.345–V.348, March 1992.

[7] T. W. Anderson. "asymptotic theory for principal component analysis". *Ann. J. Math. Stat.*, 34:122–148, 1963.

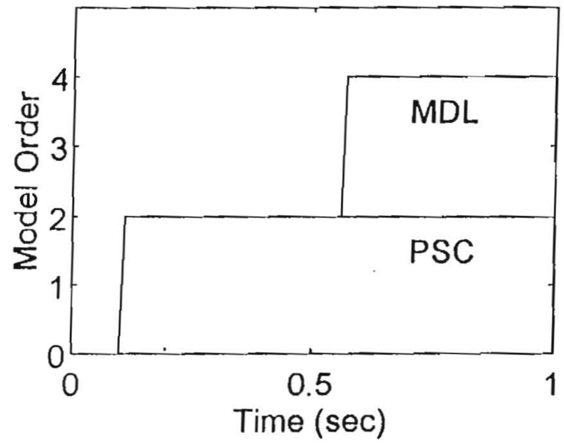

Figure 1: The detected model when the phase of the first signal varies at $t = 570$ ms.

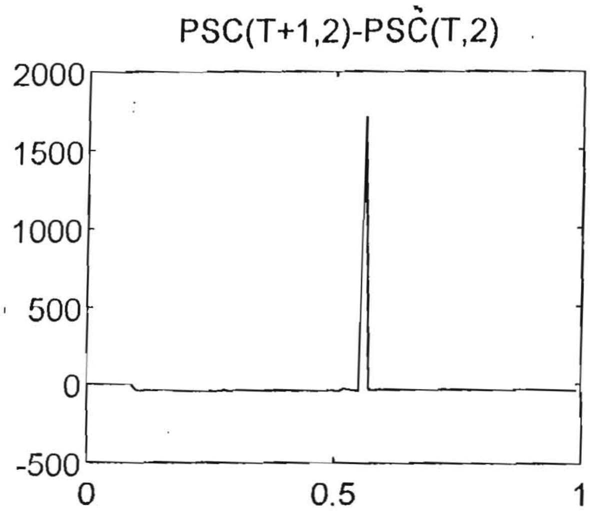

Figure 2: The difference between PSC terms for the model of order 2.

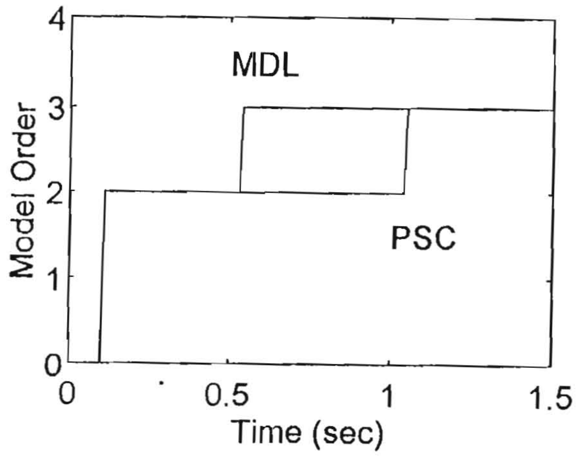

Figure 3: The detected model when one of the frequencies is time varying.