# Sinusoidal Signal Detection using the Minimum Description Length and the Predictive Stochastic Complexity

*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

*Two techniques based on the minimum description length (MDL) and the predictive stochastic complexity (PSC) are proposed for sinusoidal signal detection. The MDL and PSC criteria are the codelength of the observation and the model. The proposed techniques decompose the observation vector into its components in the signal and noise subspaces. The noise component is encoded for several model orders. The best model is selected by minimizing the codelength.*

## 1  Introduction

Sinusoidal signal detection is discussed in various fields ranging from telecommunications to array processing and spectrum estimation. Various techniques have been proposed in the literature for sinusoidal signal detection and enumeration; see [1].

Here, we propose two enumeration techniques based on the minimum description lenght (MDL) [2] and the predictive stochastic complexity (PSC) [3] principles. MDL and PSC estimate the model order by minimizing the Kullback-Leibler distance between the true model and the estimated one.

Due to temporal coherency of sinusoids, direct application of MDL and PSC generates erroneous results — the number of signals is always detected as 1. Here, we introduce an alternative approach, similar to the ones presented in [4] and [5]. 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, the noise components are encoded. This procedure is performed for all possible models and the minimum codelength is selected to estimate the number of sinusoids. The simulation study shows that the PSC has a better performancce in nonstationary environments.

## 2  Problem Formulation

Consider a time series modeled 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 $\sigma^2$. 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 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

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

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

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

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

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

where $\Omega^K = (\omega_1, \ldots, \omega_K)$ is the vector of all frequencies of the sinusoids. Using this notation,

$$\mathbf{x}(t) = \mathbf{A}(\Omega^K)\mathbf{s}(t, \Theta) + \mathbf{n}(t) \qquad (7)$$

where $\Theta = (\alpha_1, \omega_1, \phi_1, \ldots, \alpha_K, \omega_K, \phi_K)$ is the parameter vector of sinusoids. The *signal subspace* is defined as the span of $\mathbf{A}(\Omega^K)$. 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 formulate two information theoretic methods to estimate the number of signals $K$ and their frequencies $\omega_k, k = 1, \ldots, K$.

# 3   The MDL and PSC Principles

The *minimum description length* (MDL) [2] and the *predictive stochastic complexity* (PSC) [3] techniques 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

$$\mathrm{MDL}(T, k) = -\log f\left(\mathbf{X}(T)|\hat{\Psi}_T^k\right) + \frac{\nu(k)}{2}\log T \quad (8)$$

where $f(\mathbf{X}|\Psi)$ is the conditional probability density function, $\hat{\Psi}_T^k$ is the *maximum likelihood* (ML) estimate of the parameter vector $\Psi^k$ using the observations up to time $T$, and $\nu(k)$ is the number of freely chosen parameters. The model order at time $T$ is determined from

$$\hat{K} = \min_k \mathrm{MDL}(T, k) \qquad (9)$$

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 of order $k$, it amounts to

$$\mathrm{PSC}(T, k) = -\sum_{t=1}^{T} \log f\left(\mathbf{x}(t)|\hat{\Psi}_{t-1}^k\right) \qquad (10)$$

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

$$\hat{K} = \min_k \mathrm{PSC}(T, k) \qquad (11)$$

with the minimization performed over all possible models.

# 4   Signal Enumeration

In a straightforward approach, the conditional probability density of $\mathbf{X}(T)$ is determined and used in (8) and (10). 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 [4] [5]. We propose decomposing the observation vectors into their components in the signal and noise subspaces.

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

$$\mathbf{P}_s(\Omega^K) = \mathbf{A}(\Omega^K)\left(\mathbf{A}^H(\Omega^K)\mathbf{A}(\Omega^K)\right)^{-1}\mathbf{A}^H(\Omega^K). \tag{12}$$

The projection matrix onto the noise subspace is

$$\mathbf{P}_n(\Omega^K) = \mathbf{I} - \mathbf{P}_s(\Omega^K), \qquad (13)$$

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

$$\mathbf{x}(t) = \mathbf{P}_s(\Omega^K)\mathbf{x}(t) + \mathbf{P}_n(\Omega^K)\mathbf{x}(t). \qquad (14)$$

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

The formulation (14) reveals an interesting property. The information about the parameter vector $\Omega^K$, and hence the dimensionality of the signal and noise subspaces, is contained in both $\mathbf{P}_s(\Omega^K)\mathbf{x}(t)$ and $\mathbf{P}_n(\Omega^K)\mathbf{x}(t)$ vectors. One might then be able to use only one of these vectors to estimate the number of signals. In the sequel, we propose two techniques based on applying the information theoretic techniques to the noise vector only.

Let $\mathbf{x}_n(t)$ be the projection of the observation vector $\mathbf{x}(t)$ onto the noise subspace, i.e.,

$$\mathbf{x}_n(t) = \mathbf{P}_n(\Omega^K)\mathbf{x}(t). \qquad (15)$$

If $\Omega^K$ is the true parameter vector, $\mathbf{x}_n(t)$ is a zero-mean white Gaussian noise vector. The probability density function of $\mathbf{x}_n(t)$ is then

$$f(\mathbf{x}_n(t)|\Omega^K, \sigma^2) = |\pi\sigma^2\mathbf{I}|^{-1}\exp\left\{-\frac{1}{\sigma^2}\|\mathbf{x}_n(t)\|^2\right\} \tag{16}$$

where $\mathbf{I}$ is a $(M - 2K) \times (M - 2K)$ unity matrix. Define $\mathbf{X}_n(T) = [\mathbf{x}_n(t)]$, $t = 1, \ldots, T$, the projection of the observation matrix onto the noise subspace. The probability density function for $\mathbf{X}_n(T)$ is

$$f(\mathbf{X}_n(T)|\Omega^K, \sigma^2) = |\pi\sigma^2\mathbf{I}|^{-T}\exp\left\{\frac{-T}{\sigma^2}\mathrm{tr}[\mathbf{P}_n(\Omega^K)\bar{\mathbf{R}}_x]\right\} \tag{17}$$

where tr$[.]$ is the trace operator and

$$\bar{\mathbf{R}}_x = \frac{1}{T}\sum_{t=1}^{T}\mathbf{x}(t)\mathbf{x}^H(t) \qquad (18)$$

is the sample correlation matrix of the observation vector. Let us define

$$\bar{\mathbf{R}}_n(\Omega^K) = \mathbf{P}_n(\Omega^K)\bar{\mathbf{R}}_x. \qquad (19)$$

The log-likelihood function for $\mathbf{X}_n(T)$ is then

$$-\log f(\mathbf{X}_n(T)|\Omega^K,\sigma^2) = T(M-2K)\log(\sigma^2\pi)$$
$$+\frac{T}{\sigma^2}\text{tr}[\bar{\mathbf{R}}_n(\Omega^K)]. \quad (20)$$

As discussed earlier, the ML estimate of the parameter vector is used in MDL and PSC. Using the results of [6], the ML estimate of the noise power is given by

$$\hat{\sigma}^2 = \frac{1}{M-2K}\text{tr}[\bar{\mathbf{R}}_n(\Omega^K)]. \qquad (21)$$

The log-likelihood function is then

$$-\log f(\mathbf{X}_n(T)|\hat{\Omega}^K) = T(M-2K)\log\left(\frac{\pi\text{tr}[\bar{\mathbf{R}}_n(\hat{\Omega}^K)]}{M-2K}\right)$$
$$+T(M-2K). \quad (22)$$

where $\hat{\Omega}^K$ is the ML estimate of the parameter vector $\Omega^K$. Note that the number of freely chosen parameters is $K$. The MDL criterion for a model of $k$ signals is then

$$\text{MDL}(T,k) = T(M-2k)\log\left(\frac{\pi\text{tr}[\bar{\mathbf{R}}_n(\hat{\Omega}^k)]}{M-2k}\right)$$
$$+k(\frac{1}{2}\log T - 2T) \qquad (23)$$

where the constant values have not been considered in the MDL computations.

The PSC criterion is computed for all $t$ inside the window $[0,T]$. Define the sample correlation matrix at time instant $t$ by

$$\bar{\mathbf{R}}_{x_t} = \frac{1}{t}\sum_{i=1}^{t}\mathbf{x}(i)\mathbf{x}^H(i). \qquad (24)$$

The projection of this matrix onto the noise subspace for the true model is defined as

$$\bar{\mathbf{R}}_{n_t}(\hat{\Omega}_t^K) = \mathbf{P}_n(\hat{\Omega}_t^K)\bar{\mathbf{R}}_{x_t}\mathbf{P}_n(\hat{\Omega}_t^K), \qquad (25)$$

where $\hat{\Omega}_t^K$ is the ML estimate of parameter vector for true model using the observations upto time $t$.

Using the results of [6], the noise power is estimated as

$$\hat{\sigma}^2 = \frac{1}{M-2K}\text{tr}[\bar{\mathbf{R}}_{n_{t-1}}(\hat{\Omega}_{t-1}^K)]. \qquad (26)$$

Thus the PSC criterion for a model of order $k$ is

$$\text{PSC}(T,k) = (M-2k)\sum_{t=1}^{T}\log\left(\frac{\text{tr}[\bar{\mathbf{R}}_{n_t}(\hat{\Omega}_{t-1}^k)]}{M-2k}\right). \qquad (27)$$

For each criterion, an estimate of $K$ is obtained by minimizing the PSC or MDL value over all $k \leq M/2$.

# 5 Simulation Results

We include here the results for the simulation study. To reduce the computational complexity, we choose to use a root MUSIC technique to estimate $\Omega^k$.

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 1 second and decomposed into 100 non-overlapping snapshots of length 10 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 $-15$ dB to 5 dB.

Example 2: We simulate a case in which the phase of the second signal suddenly changes to $\frac{2\pi}{3}$ at $t = 570$ ms. The signals are such as in Example 1. The PSC criterion is illustrated in Fig. 1. Note that PSC depicts a break point at the location of the change. As a comparison the MDL criterion is shown in Fig. 2. As seen MDL has a jump at the change point location which might cause a wrong order detection. Fig. 3 illustrates the difference between the PSC terms. At $t = 570$ we notice an impluse in the PSC criterion which indicates that the statistics of the model has been changed. Thus, PSC can be used for change-point detection. MDL does not see this change — it is only calculated at the end of the observation window.

Example 3: In this example, 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. 4. 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 is more appropriate for a nonstationary environment.

| Noise Power (dB) | Selected Model | | | | | |
| | PSC | | | MDL | | |
| | 1 | 2 | 3 | 1 | 2 | 3 |
| --- | --- | --- | --- | --- | --- | --- |
| −15 | 0 | 100 | 0 | 0 | 100 | 0 |
| −10 | 0 | 100 | 0 | 0 | 92 | 8 |
| −5 | 0 | 100 | 0 | 0 | 0 | 100 |
| 0 | 0 | 22 | 78 | 0 | 0 | 100 |
| 5 | 0 | 0 | 100 | 0 | 0 | 100 |

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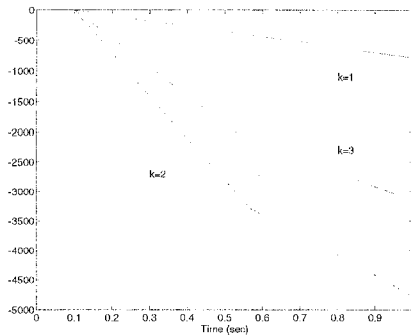


Figure 1: The PSC criterion when the phase of the first signal varies at $t = 570$ ms.

# References

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

[2] J. Rissanen. Modeling by shortest data description. *Automatica*, 14:465–471, 1979.

[3] J. Rissanen. Stochastic complexity and modeling. *Annals of Statistics*, 14:1080–1100, 1986.

[4] 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.

[5] 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.

[6] T. W. Anderson. Asymptotic theory for principal component analysis. *Ann. J. Math. Stat.*, 34:122–148, 1963.
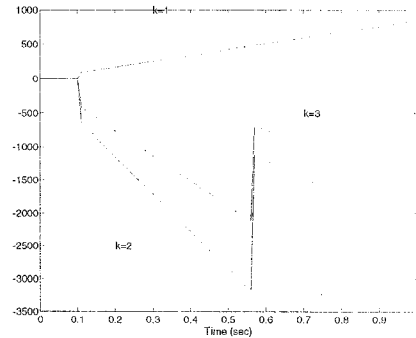
Figure 2: The MDL criterion when the phase of the first signal varies at $t = 570$ ms.
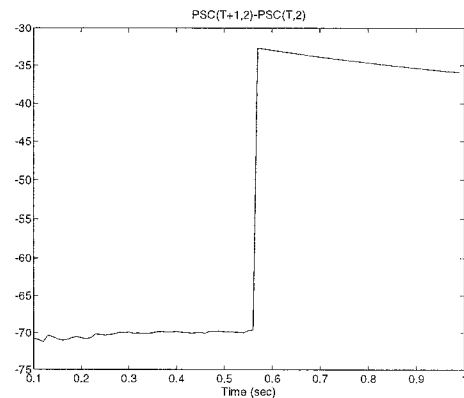


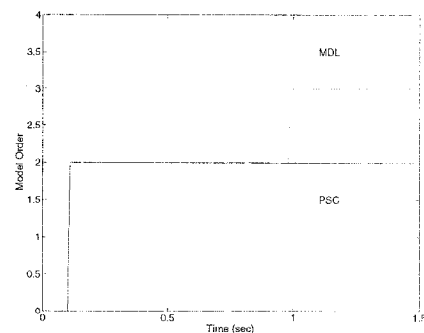Figure 3: The difference between PSC terms for the model of order 2.



Figure 4: The detected model when one of the frequencies is time varying.