STATE UNIVERSITY OF NEW YORK
AT STONY BROOK

Technical Report 634
August 24, 1992

Detection and Localization of Multiple Sources via Bayesian Predictive Densities

by

Chao-Ming Cho and *Petar M. Djuric*

Department of Electrical Engineering

* This work was supported by the National Science Foundation under Award No. MIP-9110628.

Detection and Localization of Multiple Sources via Bayesian Predictive Densities

Chao-Ming Cho and *Petar M. Djurić*

August 24, 1992

Abstract

A new approach based on Bayesian predictive density and subspace decomposition is proposed to simultaneously detect and estimate coherent and noncoherent sources. As expected, the Bayesian estimator for the directional parameters coincides with the unconditional maximum likelihood estimator when Jeffreys' noninformative priors are used. The proposed detection criterion is strongly consistent and outperforms the MDL and AIC criteria, especially in a small number of sensors and/or snapshots, and/or low SNR, without costing extra computational complexity. Simulation results demonstrating its superiority are included.

I. Introduction

In the area of array processing the most popular approaches for the detection of number of sources are based on the Akaike's information criterion (AIC) [1] and the minimum description length (MDL) principle [2, 3]. For noncoherent sources, the number of sources is determined from the "multiplicity" of the smallest eigenvalue of the sample covariance matrix without estimating the directional parameters [4] $\sim$ [6]. When the signals are coherent, this approach is not applicable since the rank of the signal covariance matrix is reduced. To effectively solve this problem, Wax has proposed a subspace decomposition approach, based on above information criteria, to detect and estimate coherent sources [7]. However, the AIC criterion suffers two drawbacks. It tends to asymptotically overestimate the number of sources and its probability of error can not reach zero even at a high signal-to-noise ratio (SNR). On the other hand, the MDL criterion is consistent, but it overemphasizes the performance when a larger number of snapshots are available, sacrificing the performance at low SNR and/or a small number of snapshots. Unfortunately, in the most frequent cases, it might be that the energy of the signals impinging on the array is low and the number of snapshots is limited.

There is another question of interest. Assume that the number of snapshots is fixed and the number of sensor is decreased, the relative penalty function of any information criterion is supposed to be decreased, otherwise it may induce a underestimated result when the SNR is low. Someone may doubt whether the penalty terms of the AIC and MDL criteria do reflect this situation or not. In this paper, we are not try to analyze the effect of decreasing the number of sensors. However, the penalty term of the information criterion is expected as the function of the number of sensors.

In our paper, a new criterion is proposed to simultaneously detect the number of coherent or noncoherent sources and estimate their directional parameters. The solution is obtained by using Bayesian predictive densities (BPD) [8] and subspace decomposition [7]. When compared to the AIC and MDL criteria in [7], the relative penalty term of the proposed method is greater than that of the AIC and smaller than that of the MDL. Furthermore, it is also the function of the number of sensors. This exact penalty term leads to better detection performance, especially in the cases that the number of snapshots is small and/or the SNR is low, without increasing computational complexity. Unlike the AIC criterion, the proposed criterion is strongly consistent. As expected, the Bayesian estimator coincides with the unconditional maximum likelihood (ML) estimator [9, 10] which has been shown to be consistent and efficient.

The paper is organized as follows. In Section II we formulate the problem and declare the assumptions. The Bayesian predictive density criterion is derived in Section III. In Section IV some simulation results are performed to demonstrate the improved performance of the proposed approach when compared to the MDL and AIC œ[250zcriteria. Finally, the conclusion is given in Section V.

II. Problem Formulation

Consider that q superimposed far-field sources are measured by N sensors, and assume $q < N$. The locations and the directional characteristics of the sensors are allowed to be arbitrary. The sources emit narrow-band wavefronts centered at a known frequency, ω_0, and impinge on the sensors in a planar manner.

The observed data at the pth sensor and the ith snapshot is expressed, by the complex envelope representation, as

$$y_{i,p} = \sum_{l=1}^{q} s_{i,l} e^{jp\theta_l} + n_{i,p}, \quad i = 1, 2, \ldots, M \tag{1}$$

where θ_l is the directional parameter (direction-of-arrival) of the lth source, assumed distinct from the other sources, $s_{i,l}$ is the complex amplitude of the lth source as received at reference point, and M denotes the number of snapshots. The $n_{i,p}$ is the additive complex noise at the pth sensor.

The model for the ith snapshot can be compactly described by the following vector notations:

$$\mathbf{y}_i = \mathbf{D}(\boldsymbol{\theta}_{(q)})\mathbf{s}_i + \mathbf{n}_i, \quad i = 1, 2, \ldots, M \tag{2}$$

with

$$\mathbf{y}_i = [\, y_{i,1} \; y_{i,2} \cdots y_{i,N} \,]^T \tag{3.a}$$

$$\mathbf{D}(\boldsymbol{\theta}_{(q)}) = [\, \mathbf{d}(\theta_1) \; \mathbf{d}(\theta_2) \ldots \mathbf{d}(\theta_q) \,] \tag{3.b}$$

$$\mathbf{d}(\theta_l) = [\, 1 \; e^{j\theta_l} \ldots e^{j(N-1)\theta_l} \,]^T \tag{3.c}$$

$$\mathbf{s}_i = [s_{1,i}\ s_{2,i} \dots s_{q,i}]^T \qquad (3.d)$$

$$\mathbf{n}_i = [n_{i,1}\ n_{i,2} \dots n_{i,N}]^T \qquad (3.e)$$

where $\mathbf{D}(\boldsymbol{\theta}_{(q)})$ is an $N \times q$ Vandermonde matrix consisting of q steering vectors $\mathbf{d}(\theta_l)$'s. Any q distinct steering vectors from the array manifold are linearly independent. $\mathbf{y}_i$ is an $N \times 1$ observed data vector, $\mathbf{s}_i$ is an $q \times 1$ signal vector, and $\mathbf{n}_i$ is an $N \times 1$ noise vectors. T denotes transpose operation.

We assume that the signal sample vectors $\mathbf{s}_i$'s are statistically independent and identical complex Gaussian random vectors with zero mean and an unknown covariance matrix. The noise in each sensor is a stationary, ergodic, complex Gaussian process with zero mean and an unknown variance. The noise samples are also assumed to be uncorrelated from sensor to sensor and from the impinging signals. The covariance matrix of the observed data is then given by

$$\boldsymbol{\Sigma}_{yy} = \mathbf{D}(\boldsymbol{\theta}_{(q)})\mathbf{R}_{ss}\mathbf{D}^H(\boldsymbol{\theta}_{(q)}) + \mathbf{R}_{nn} \qquad (4)$$

where $\mathbf{R}_{ss}$ is an $q \times q$ unknown signal covariance matrix. The signals may be uncorrelated (noncoherent), partially correlated, or fully correlated (coherent). When the signals are coherent, one signal might be a scaled and delayed version of the other, especially in multipath propagation. $\mathbf{R}_{nn}$ denotes an $N \times N$ unknown noise covariance matrix, and in our case it is equal to $\sigma_n^2\mathbf{I}$. H denotes conjugation and transposition.

When the above assumptions hold, the problem can be stated as follows. Given the observed data samples, we desire to simultaneously detect the number of the coherent or noncoherent sources and estimate their directional parameters.

III. Bayesian Predictive Density Approach

In the sequel a method based on Bayesian inference techniques is proposed to solve the above problem. Let $\mathcal{H}_k$ denotes the hypothesis that k sources are present, and $k \in \{0, \dots, N-1\}$. Under the hypothesis $\mathcal{H}_k$, the model is given by

$$\mathbf{y}_i = \mathbf{D}(\boldsymbol{\theta}_{(k)})\mathbf{s}_i + \mathbf{n}_i, \quad i = 1, 2, \dots, M \qquad (5)$$

The estimates of q and $\boldsymbol{\theta}$, denoted by $\hat{q}$ and $\hat{\boldsymbol{\theta}}$, are determined from the cost function [11],

$$\hat{q},\, \hat{\boldsymbol{\theta}} = \arg \min_{k\in\{0,\ldots,N-1\};\theta\in\Theta} \{ -\log f(\boldsymbol{\theta},\, \mathcal{H}_k \,|\, \mathbf{y}) \} \tag{6}$$

where Θ is the "field of view," and $f(\boldsymbol{\theta},\, \mathcal{H}_k \,|\, \mathbf{y})$ is the posterior distribution of $\mathcal{H}_k$ and $\boldsymbol{\theta}$. By using Bayes' rule, the posterior distribution of $\mathcal{H}_k$ and $\boldsymbol{\theta}$ can be obtained by

$$f(\boldsymbol{\theta},\, \mathcal{H}_k \,|\, \mathbf{y}) = \frac{f(\mathbf{y} \,|\, \boldsymbol{\theta},\, \mathcal{H}_k)}{f(\mathbf{y})} f(\boldsymbol{\theta} \,|\, \mathcal{H}_k) f(\mathcal{H}_k) \tag{7}$$

where $f(\mathbf{y} \,|\, \boldsymbol{\theta},\, \mathcal{H}_k)$ is the likelihood function (LF) of $\mathcal{H}_k$ and $\boldsymbol{\theta}$, $f(\boldsymbol{\theta})$ is the *a priori* distribution of $\boldsymbol{\theta}$, $f(\mathcal{H}_k)$ is the probability of the model $\mathcal{H}_k$, and $f(\mathbf{y})$ is the marginal distribution of $\mathbf{y}$. When the probabilities of all hypotheses $\mathcal{H}_k$ are equal, the criterion (6) amounts to maximization of $f(\mathbf{y} \,|\, \boldsymbol{\theta},\, \mathcal{H}_k) f(\boldsymbol{\theta} \,|\, \mathcal{H}_k)$.

Within the framework of Bayesian theory, the LF $f(\mathbf{y} \,|\, \boldsymbol{\theta},\, \mathcal{H}_k)$ might be obtained by integrating out any unwanted or "nuisance" parameters, e.g., $\mathbf{R}_{ss}$ and σ_n, in the completed LF [12] $\sim$ [14]. Before this marginalization, the prior distributions of nuisance parameters have to be carefully chosen. Unless the priors are proper and supported by satisfactory physical or logical arguments, we prefer to use noninformative priors because they reflect the ignorance of the nuisance parameters. However, these priors may introduce unjustified model selection criteria when the LF of $\boldsymbol{\theta}$, $\mathcal{H}_k$ is not normalized [8]. To circumvent this deficiency, we propose a Bayesian predictive density (BPD) criterion to estimate q and $\boldsymbol{\theta}$. The BPD criterion is given by

$$\hat{q},\, \hat{\boldsymbol{\theta}} = \arg \min_{k,\,\theta} \{ -\log f(\boldsymbol{\xi}_2 \,|\, \boldsymbol{\xi}_1,\, \boldsymbol{\theta},\, \mathcal{H}_k) - \log f(\boldsymbol{\theta} \,|\, \mathcal{H}_k) \} \tag{8}$$

where $\boldsymbol{\xi}_1 = \{ \mathbf{y}_1, \ldots, \mathbf{y}_L \}$, $\boldsymbol{\xi}_2 = \{ \mathbf{y}_{L+1}, \ldots, \mathbf{y}_M \}$, and $1 < L < M$. The selection of L will be discussed later. The function $f(\boldsymbol{\xi}_2 \,|\, \boldsymbol{\xi}_1,\, \boldsymbol{\theta},\, \mathcal{H}_k)$ is called a Bayesian predictive density of $\boldsymbol{\xi}_2$ according to $\boldsymbol{\xi}_1$, the parameter vector $\boldsymbol{\theta}$, and the hypothesis $\mathcal{H}_k$. Using the Bayesian approach, the BPD function can be expressed as

$$f(\boldsymbol{\xi}_2 \,|\, \boldsymbol{\xi}_1,\, \boldsymbol{\theta},\, \mathcal{H}_k) \;=\; \frac{f(\mathbf{y}_{(M)} \,|\, \boldsymbol{\theta},\, \mathcal{H}_k)}{f(\mathbf{y}_{(L)} \,|\, \boldsymbol{\theta},\, \mathcal{H}_k)}$$

$$= \frac{\int_\Psi f(\mathbf{y}_{(M)} \,|\, \Psi,\, \boldsymbol{\theta},\, \mathcal{H}_k\,) f(\,\Psi \,|\, \boldsymbol{\theta},\, \mathcal{H}_k\,)\, d\Psi}{\int_\Psi f(\mathbf{y}_{(L)} \,|\, \Psi,\, \boldsymbol{\theta},\, \mathcal{H}_k\,) f(\,\Psi \,|\, \boldsymbol{\theta},\, \mathcal{H}_k\,)\, d\Psi} \tag{9}$$

where $\mathbf{y}_{(L)} = \{\boldsymbol{\xi}_1\}$, and $\mathbf{y}_{(M)} = \{\boldsymbol{\xi}_1,\, \boldsymbol{\xi}_2\}$, and the nuisance parameters $\Psi = \{\mathbf{R}_{ss},\, \mathbf{R}_{nn}\}$. $f(\,\Psi \,|\, \boldsymbol{\theta},\, \mathcal{H}_k\,)$ denotes the priors of the nuisance parameters under the conditions which $\boldsymbol{\theta}$ and $\mathcal{H}_k$ are given.

To properly reflect our ignorance of $\boldsymbol{\theta}$, the prior distribution $f(\,\boldsymbol{\theta} \,|\, \mathcal{H}_k\,)$ is selected to be an uniform distribution in the search domain of $\boldsymbol{\theta}$. Since the degrees of freedom in the search domain equal the number of sample points in one experiment, we choose the noninformative prior according to

$$f(\,\boldsymbol{\theta} \,|\, \mathcal{H}_k\,) = \frac{(N-k)!k!}{N!} \tag{10}$$

Note that the prior of $\boldsymbol{\theta}$ is independent of the number of experiments, M. When the number of samples is large, the prior is negligible.

In order to deal with the marginalizations of the nuisance parameters of signals and noises separately, the observed data space may be split into the two complementary subspaces [7]. The subspace, which is spanned by the column of the matrix $\mathbf{D}(\boldsymbol{\theta}_{(k)})$, is referred to as the *signal subspace.* Another subspace, which is the orthogonal space to the signal subspace, is called the *noise subspace.* According to this subspace decomposition approach, the observed data vector $\mathbf{y}$ is then decomposed into two subspace vectors by

$$\mathbf{y} = \mathbf{G}(\boldsymbol{\theta}_{(k)}) \begin{bmatrix} \mathbf{x}_s \\ \mathbf{x}_n \end{bmatrix} \tag{11}$$

where $\mathbf{x}_s$ denotes the $k \times 1$ signal subspace vector, and $\mathbf{x}_n$ denotes the $(N-k) \times 1$ noise subspace vector. $\mathbf{G}(\boldsymbol{\theta}_{(k)})$ denotes an $N \times N$ unitary coordinate transformation matrix and satisfies the following identities,

$$\mathbf{P}(\boldsymbol{\theta}_{(k)})\mathbf{y} = \mathbf{G}(\boldsymbol{\theta}_{(k)}) \begin{bmatrix} \mathbf{x}_s \\ \mathbf{0} \end{bmatrix} \tag{12}$$

and

$$\mathbf{P}^{\perp}(\boldsymbol{\theta}_{(k)})\mathbf{y} = \mathbf{G}(\boldsymbol{\theta}_{(k)}) \begin{bmatrix} \mathbf{0} \\ \mathbf{x}_n \end{bmatrix} \tag{13}$$

where $\mathbf{P}(\boldsymbol{\theta}_{(k)})$ and $\mathbf{P}^{\perp}(\boldsymbol{\theta}_{(k)})$ are two complementary projection matrices. The matrix $\mathbf{P}(\boldsymbol{\theta}_{(k)})$ is given by

$$\mathbf{P}(\boldsymbol{\theta}_{(k)}) = \mathbf{D}(\boldsymbol{\theta}_{(k)})(\,\mathbf{D}^H(\boldsymbol{\theta}_{(k)})\mathbf{D}(\boldsymbol{\theta}_{(k)})\,)^{-1}\mathbf{D}^H(\boldsymbol{\theta}_{(k)}) \tag{14}$$

and it projects onto the signal subspace, while the projection matrix $\mathbf{P}^{\perp}(\boldsymbol{\theta}_{(k)})$ is obtained from

$$\mathbf{P}^{\perp}(\boldsymbol{\theta}_{(k)}) = \mathbf{I} - \mathbf{P}(\boldsymbol{\theta}_{(k)}) \tag{15}$$

and it projects onto the noise subspace.

According to the linear transformation (11), the LF $f(\,\mathbf{y}\,|\,\Psi,\,\boldsymbol{\theta},\,\mathcal{H}_k\,)$ can be rewritten as

$$f(\,\mathbf{y}\,|\,\Psi,\,\boldsymbol{\theta},\,\mathcal{H}_k\,) = \frac{1}{|\mathcal{J}(\mathbf{x}_s,\mathbf{x}_n)|} f(\,\mathbf{x}_s,\mathbf{x}_n\,|\,\Psi,\,\boldsymbol{\theta},\,\mathcal{H}_k\,) \tag{16}$$

where $\mathcal{J}(\mathbf{x}_s,\mathbf{x}_n)$ is the Jacobian of the transformation (11). $|\cdot|$ denotes absolute value or modulus. This Jacobian is

$$|\mathcal{J}(\mathbf{x}_s,\mathbf{x}_n)| = |\,\mathsf{det}(\mathbf{G}^H(\boldsymbol{\theta}_{(k)}))\,| = 1 \tag{17}$$

Assume that the signal and noise subspace vectors are independent and uncorrelated. Without loss of any information, the nuisance parameters Ψ are also split into two nuisance parameter sets, Ψ_s and Ψ_n, that belong to complementary subspaces. The parameter set Ψ_s depends only on the signal subspace vector. In contrast, the set Ψ_n represents the nuisance parameters of the noise subspace components. Therefore, (16) can be rewritten as

$$f(\,\mathbf{y}\,|\,\Psi,\,\boldsymbol{\theta},\,\mathcal{H}_k\,) = f(\,\mathbf{x}_s\,|\,\Psi_s,\,\boldsymbol{\theta},\,\mathcal{H}_k\,)f(\,\mathbf{x}_n\,|\,\Psi_n,\,\boldsymbol{\theta},\,\mathcal{H}_k\,) \tag{18}$$

where $\Psi_s = \mathbf{\Sigma}_{ss}$ and $\Psi_n = \mathbf{\Sigma}_{nn}$. The numerator of (9) can be modified as

$$\begin{aligned}
&\int_{\Psi} f(\mathbf{y}_{(M)} \mid \Psi, \boldsymbol{\theta}, \mathcal{H}_k) f(\Psi \mid \boldsymbol{\theta}, \mathcal{H}_k)\, d\Psi \\
&= \int_{\Psi_s} f(\mathbf{x}_{s,(M)} \mid \Psi_s, \boldsymbol{\theta}, \mathcal{H}_k) f(\Psi_s \mid \boldsymbol{\theta}, \mathcal{H}_k)\, d\Psi_s \\
&\times \int_{\Psi_n} f(\mathbf{x}_{n,(M)} \mid \Psi_n, \boldsymbol{\theta}, \mathcal{H}_k) f(\Psi_n \mid \boldsymbol{\theta}, \mathcal{H}_k)\, d\Psi_n \\
&= f(\mathbf{x}_{s,(M)} \mid \boldsymbol{\theta}, \mathcal{H}_k) f(\mathbf{x}_{n,(M)} \mid \boldsymbol{\theta}, \mathcal{H}_k) \qquad (19)
\end{aligned}$$

The dominator of (9) is manipulated in the same way.

From the assumptions in section II and (11), it follows that the unknown signal subspace vector $\mathbf{x}_s$ is modeled by a complex Gaussian process with zero mean and an unknown covariance matrix $\mathbf{\Sigma}_{ss}$. Given the information of $\mathbf{\Sigma}_{ss}$ and $\boldsymbol{\theta}$, the distribution of $\mathbf{x}_s$ is

$$f(\mathbf{x}_{s,(M)} \mid \mathbf{\Sigma}_{ss}^{-1}, \boldsymbol{\theta}, \mathcal{H}_k) = (\frac{1}{\pi})^{kM} [\mathsf{det}(\mathbf{\Sigma}_{ss}^{-1})]^M \exp\{ -\sum_{i=1}^{M} \mathbf{x}_{s,i}^H \mathbf{\Sigma}_{ss}^{-1} \mathbf{x}_{s,i} \} \qquad (20)$$

where $\mathsf{det}(\cdot)$ denotes the determinant. For the white noise model, $\mathbf{x}_n$ is an $(N-k)\times 1$ complex Gaussian random vector with zero mean and covariance matrix $\mathbf{\Sigma}_{nn} = \sigma_n^2 \mathbf{I}$, i.e.,

$$f(\mathbf{x}_{n,(M)} \mid \sigma_n, \boldsymbol{\theta}, \mathcal{H}_k) = \frac{1}{(\pi\sigma_n^2)^{(N-k)M}} \exp\{ -\frac{1}{\sigma_n^2} \sum_{i=1}^{M} \mathbf{x}_{n,i}^H \mathbf{x}_{n,i} \} \qquad (21)$$

Assume that the priors of Ψ_s and Ψ_n are independent of $\boldsymbol{\theta}$, we get

$$f(\Psi_s \mid \boldsymbol{\theta}, \mathcal{H}_k) = f(\mathbf{\Sigma}_{ss}^{-1} \mid \mathcal{H}_k) \qquad (22)$$

$$f(\Psi_n \mid \boldsymbol{\theta}, \mathcal{H}_k) = f(\sigma_n \mid \mathcal{H}_k) \qquad (23)$$

Since we have no information about $\mathbf{\Sigma}_{ss}^{-1}$ in (22), we may choose the noninformative prior distribution for $f(\mathbf{\Sigma}_{ss}^{-1})$ using Jeffreys' invariance theory [12]. In Jeffreys' invariance theory, the noninformative prior distribution for a set of parameters is taken to be proportional to the square root of the determinant of the Fisher's information matrix. Hence, the prior distribution of $\mathbf{\Sigma}_{ss}^{-1}$ can be shown to be [16]

$$\begin{aligned}
f(\mathbf{\Sigma}_{ss}^{-1} \mid \mathcal{H}_k) &\propto [\mathsf{det}(\frac{\partial \mathbf{\Sigma}_{ss}}{\partial \mathbf{\Sigma}_{ss}^{-1}})]^{\frac{1}{2}} \\
&\propto [\mathsf{det}(\mathbf{\Sigma}_{ss}^{-1})]^{-k} \qquad (24)
\end{aligned}$$

where $|\frac{\partial \mathbf{\Sigma}_{ss}}{\partial \mathbf{\Sigma}_{ss}^{-1}}|$ is the Jacobian of the information from the elements σ_{ij} of $\mathbf{\Sigma}_{ss}$ to the elements σ^{ij} of $\mathbf{\Sigma}_{ss}^{-1}$. Like (24), we have to select a noninformative prior distribution of σ_n^2 to reflect our ignorance of the noise variance. This noninformative prior is given by [12]

$$f(\,\sigma_n\,|\,\mathcal{H}_k\,) \propto \frac{1}{\sigma_n} \tag{25}$$

Thus, the LF of the signal subspace vector in (18) is rewritten as

$$\begin{aligned} & f(\,\mathbf{x}_{s,(M)}\,|\,\boldsymbol{\theta},\,\mathcal{H}_k\,) \\ = & \int_{\mathbf{\Sigma}_{ss}^{-1}} (\frac{1}{\pi})^{kM} [\mathsf{det}(\mathbf{\Sigma}_{ss}^{-1})]^M \exp\{\,-\mathsf{tr}(M\cdot\hat{\mathbf{\Sigma}}_{ss,(M)}\mathbf{\Sigma}_{ss}^{-1})\,\}\cdot[\mathsf{det}(\mathbf{\Sigma}_{ss}^{-1})]^{-k} d\mathbf{\Sigma}_{ss}^{-1} \end{aligned} \tag{26}$$

where tr denotes the trace of the matrix, and

$$\hat{\mathbf{\Sigma}}_{ss,(M)} = \frac{1}{M}\sum_{i=1}^{M} \mathbf{x}_{s,i}\,\mathbf{x}_{s,i}^H \tag{27}$$

The estimator $\hat{\mathbf{\Sigma}}_{ss}$ is a sufficient statistic for the Hermitian covariance matrix $\mathbf{\Sigma}_{ss}$. The integration of (26) can be manipulated by changing variables in a complex Wishart distribution [17], then

$$f(\,\mathbf{x}_{s,(M)}\,|\,\boldsymbol{\theta},\,\mathcal{H}_k\,) = M^{-Mk}[\,\mathsf{det}(\hat{\mathbf{\Sigma}}_{ss,(M)})\,]^{-M}(\frac{1}{\pi})^{kM}\pi^{k(k-1)/2}\prod_{l=0}^{k-1}\Gamma[M-l] \tag{28}$$

Next, the LF of the noise subspace vector is obtained by

$$\begin{aligned} & f(\,\mathbf{x}_{n,(M)}\,|\,\boldsymbol{\theta},\mathcal{H}_k\,) \\ = & \int_{\sigma_n} (\frac{1}{\pi})^{(N-k)M}\sigma_n^{-2(N-k)M}\exp\{\,-\frac{M}{\sigma_n^2}\mathsf{tr}(\hat{\mathbf{\Sigma}}_{nn,(M)})\,\}\,\sigma_n^{-1}\,d\sigma_n \\ = & (\frac{1}{\pi})^{(N-k)M}\frac{1}{2}\Gamma[M(N-k)]\cdot M^{-M(N-k)}\cdot[\,\mathsf{tr}(\hat{\mathbf{\Sigma}}_{nn,(M)})\,]^{-M(N-k)} \end{aligned} \tag{29}$$

where

$$\hat{\mathbf{\Sigma}}_{nn,(M)} = \frac{1}{M}\sum_{i=1}^{M} \mathbf{x}_{n,i}\,\mathbf{x}_{n,i}^H \tag{30}$$

Multipling (28) and (29), it yields

$$\begin{aligned} f(\mathbf{y}_{(M)} \,|\, \boldsymbol{\theta}, \mathcal{H}_k) &= f(\mathbf{x}_{s,(M)} \,|\, \boldsymbol{\theta}, \mathcal{H}_k) f(\mathbf{x}_{n,(M)} \,|\, \boldsymbol{\theta}, \mathcal{H}_k) \\ &= [\, C_{(M)}(\boldsymbol{\theta}_{(k)})\,]^{-M} \left(\frac{1}{M\pi}\right)^{MN} (N-k)^{-M(N-k)} \pi^{k(k-1)/2} \\ &\times \Gamma[M(N-k)] \prod_{l=0}^{k-1} \Gamma[M-l] \end{aligned} \tag{31}$$

where

$$C_{(M)}(\boldsymbol{\theta}_{(k)}) = \left(\mathsf{det}(\hat{\boldsymbol{\Sigma}}_{ss,(M)})\right) \left(\frac{1}{(N-k)} \mathsf{tr}(\hat{\boldsymbol{\Sigma}}_{nn,(M)})\right)^{(N-k)} \tag{32}$$

The data term $C_{(M)}(\boldsymbol{\theta}_{(k)})$ can be directly obtained from the observed data samples. As shown in [7], we get

$$\begin{aligned} & C_{(M)}(\boldsymbol{\theta}_{(k)}) \\ &= \mathsf{det}\left(\mathbf{P}(\boldsymbol{\theta}_{(k)}) \hat{\boldsymbol{\Sigma}}_{yy,(M)} \mathbf{P}(\boldsymbol{\theta}_{(k)}) + \frac{1}{(N-k)} \mathsf{tr}(\mathbf{P}^{\perp}(\boldsymbol{\theta}_{(k)}) \hat{\boldsymbol{\Sigma}}_{yy,(M)}) \mathbf{P}^{\perp}(\boldsymbol{\theta}_{(k)}) \right) \end{aligned} \tag{33}$$

where

$$\hat{\boldsymbol{\Sigma}}_{yy,(M)} = \frac{1}{M} \sum_{i=1}^{M} \mathbf{y}_i \, \mathbf{y}_i^H \tag{34}$$

Furthermore, $C_{(M)}(\boldsymbol{\theta}_{(k)})$ can also be computed in terms of the eigenvalues of the matrices involved. Using the well-known invariance properties of the unitary transformation, it becomes

$$C_{(M)}(\boldsymbol{\theta}_{(k)}) = \left(\prod_{i=1}^{k} \lambda_i^{(s)}(\boldsymbol{\theta}_{(k)}) \right) \left(\frac{1}{(N-k)} \sum_{i=1}^{N-k} \lambda_i^{(n)}(\boldsymbol{\theta}_{(k)}) \right)^{(N-k)} \tag{35}$$

where $\lambda_i^{(s)}(\boldsymbol{\theta}_{(k)})$'s are k nonzero eigenvalues of the rank-k matrix $\mathbf{P}(\boldsymbol{\theta}_{(k)}) \hat{\boldsymbol{\Sigma}}_{yy,(M)} \mathbf{P}(\boldsymbol{\theta}_{(k)})$ and $\lambda_i^{(n)}(\boldsymbol{\theta}_{(k)})$'s are $(N-k)$ nonzero eigenvalues of the rank-$(N-k)$ matrix $\mathbf{P}^{\perp}(\boldsymbol{\theta}_{(k)}) \hat{\boldsymbol{\Sigma}}_{yy,(M)} \mathbf{P}^{\perp}(\boldsymbol{\theta}_{(k)})$.

The dominator of (9) is derived by the same approach. It results in

$$\begin{aligned} f(\mathbf{y}_{(L)} \,|\, \boldsymbol{\theta}, \mathcal{H}_k) &= [\, C_{(L)}(\boldsymbol{\theta}_{(k)})\,]^{-L} \left(\frac{1}{L\pi}\right)^{LN} \\ &\times (N-k)^{-L(N-k)} \pi^{k(k-1)/2} \Gamma[L(N-k)] \prod_{l=0}^{k-1} \Gamma[L-l] \end{aligned} \tag{36}$$

Thus, substituting (10), (31) and (36) into (8), the BPD is obtained to be

$$
\begin{aligned}
-\log f(\boldsymbol{\xi}_2 \,|\, \boldsymbol{\xi}_1, \boldsymbol{\theta}, \mathcal{H}_k) &= M \cdot \log C_{(M)}(\boldsymbol{\theta}_{(k)}) - L \cdot \log C_{(L)}(\boldsymbol{\theta}_{(k)}) \\
&+ (M - L)(N - k)\log(N - k) + \log \frac{\Gamma[L(N-k)]}{\Gamma[M(N-k)]} \\
&+ \sum_{l=0}^{k-1} \log \frac{\Gamma[L-l]}{\Gamma[M-l]} + MN \log M\pi - LN \log L\pi \qquad (37)
\end{aligned}
$$

Note that L has to be greater than the dimension of the signal subspace vector to satisfy the minimum number of degrees of freedom in a complex Wishart distribution. Furthermore, we know that L should be selected as small as possible to minimize the information loss [8]. Therefore, we choose $L = N - 1$ for the maximum possible number of k, without contradicting the above constraints. To simplify the computation of (37), we may assume that $C_{(M)}(\boldsymbol{\theta}_{(k)})$ and $C_{(L)}(\boldsymbol{\theta}_{(k)})$ are approximately equal. Combining (10) and (37), and ignoring the terms which are independent of k and $\boldsymbol{\theta}$, the final BPD criterion is expressed as

$$
\hat{q}, \hat{\boldsymbol{\theta}} = \arg \min_{k,\theta} \{ (M - N + 1) \cdot \log C_{(M)}(\boldsymbol{\theta}_{(k)}) + T(k) \} \qquad (38)
$$

where the penalty function is

$$
\begin{aligned}
T(k) &= (M - N + 1)(N - k)\log(N - k) + \log \frac{\Gamma[(N-1)(N-k)]}{\Gamma[M(N-k)]} \\
&+ \sum_{l=0}^{k-1} \log \frac{\Gamma[N-l-1]}{\Gamma[M-l]} + \log \frac{N!}{(N-k)!k!} \qquad (39)
\end{aligned}
$$

Under the same assumptions in section II, the MDL and AIC criterion are given by [7]

$$
MDL(k) = \min_{k} \{ M \cdot \log C_{(M)}(\boldsymbol{\theta}_{(k)}) + \frac{1}{2}k(k+1)\log M \} \qquad (40)
$$

and

$$
AIC(k) = \min_{k} \{ M \cdot \log C_{(M)}(\boldsymbol{\theta}_{(k)}) + k(k+1) \} \qquad (41)
$$

Clearly, the BPD criterion has the same data term, but different penalty function.

The Bayesian estimator of $\boldsymbol{\theta}$ is obtained by maximizing $C_{(M)}(\boldsymbol{\theta}_{(k)})$ because the penalty term is independent of $\boldsymbol{\theta}$. This solution coincides with the ML estimator derived by Böhme [9], Jaffer [10], and Wax [7]. As expected, the Bayesian estimator yields the same result as the so-called stochastic ML estimator when the Jeffreys' non-informative priors are used [15]. This estimator has been shown to be asymptotically unbiased (consistent) and statistically efficient, i.e., the estimation error covariance attains the Cramer-Rao bound asymptotically [18, 19].

For the detection part, the BPD criterion yields a different penalty function from the MDL and AIC criteria. Unlike the AIC criterion, the BPD criterion is strongly consistent, such that $\lim_{M\to\infty} Pr(\hat{q} = q) = 1$ (see Appendix). When compared to the MDL's penalty function, the relative penalty term of the BPD approach is smaller, and is also a function of the number of sensors. This exact penalization leads to better detection performance, especially in the case that the number of snapshots is small and/or the SNR is low. The superiority of the proposed BPD criterion is demonstrated in the next section.

The computation of the BPD estimator based on (38) is complicated since a nonlinear and multimodal k-dimensional maximization problem has to be carried out. In order to efficiently solve this problem, the alternating maximization technique [20], or a dynamic programming approach [21] can be used. These techniques transform the k-dimensional problem into a sequence of one-dimensional searches, and this considerably reduces the computational load. In addition, some numerical approaches [22] may be applied to further improve the efficiency in computation and the convergence in the optimal search.

IV. Simulation results

To examine the performance of the BPD criterion, six simulated experiments were performed, each with 100 Monte Carlo runs. The detection performance was obtained by counting the number of correctly estimated q in 100 runs. The experiments compared the detection performance of the BPD, AIC and MDL criteria for two cases, coherent and noncoherent sources. Each case was examined in term of M, SNR, and N.

Figures 1 and 2 show the comparisons of the detection performance in term of the number of snapshots for two equal power coherent and noncoherent sources, respectively. We observe that the BPD criterion outperforms the MDL criterion in both experiments, especially when M is small and the signals are uncorrelated. Although the AIC criterion yields the better performance for small M, it became inconsistent as M increasing. In experiments 3 and 4, we compared the performance of the BPD, AIC, and MDL criteria for various SNR. In Figures 3 and 4, we observe that the BPD outperforms the MDL criterion when the SNR is small. Like experiments 1 and 2, the detection probability of the AIC criterion can not approach one even when the SNR is high. Figures 5 and 6 show the performance of three criteria in term of N when $M = 50$ and SNR = -6 dB. The results also demonstrate the superiority of the BPD approach.

These six experiments show that the gain in the detection performance of the BPD approach is larger in the cases of small M, SNR, and N. In addition, the gain is greater when the sources are noncoherent. Generally speaking, the BPD criterion has a moderate penalty function which is greater than that of the AIC and smaller than that of MDL. The AIC tends asymptotically to overestimate q, while the MDL tends to underestimate q when the SNR and the number of snapshots are small.

Since the Bayesian estimator (32) coincides with the unconditional ML estimator, its estimation performance is not examined here. This estimator has been shown that it outperforms the conditional ML estimation for the cases that the sources are uncorrelated or fully correlated.

V. Conclusions

In this paper, we have applied Bayesian inference techniques to the detection and estimation of noncoherent or coherent sources. The solution is undertaken by maximization of the posterior distribution of $\mathcal{H}_k$ and $\boldsymbol{\theta}$. To combine separate marginal distributions of the signal and noise subspace vectors, Bayesian predictive density approach provides a justified decision criterion without any complicated normalization. Since the nuisance parameters have been eliminated by marginalization, the uncertainty introduced in the estimation of the parameters is reduced. In other words, we reduce the number of the unknown parameters, which have to be estimated, and derive a more exact penalty term. When N, M, and SNR are small, this exact penalization results in the better detection performance without increasing computational complexity. Furthermore, the BPD is consistent for estimating the number of sources and the directional parameters.

In the AIC and MDL criteria, the penalty term is derived by asymptotical approach and becomes the function of number of free parameters involved. The free parameters in these approaches includes the wanted and nuisance parameters. The more parameters involved, the larger uncertainty is introduced in the decision function. In contrast, our approach reduce the number of free parameters in the cost function. The penalization for nuisance parameters is automatically generated after marginalization, without using asymptotical approach. This is the key to the improved performance.

When the number of sensors is large, the performance of the proposed method can also be improved by two ways. Since the possible number of sources may be much less than N, in this case, we suggest to use a small L instead of $N-1$ in (37). The smaller L, the less information loss in the BPD criterion. In addition, the prior of $\boldsymbol{\theta}$ can be replaced by using Jeffreys' prior. Although this prior is generated from the asymptotical assumption, it provides a good performance even when N is not large.

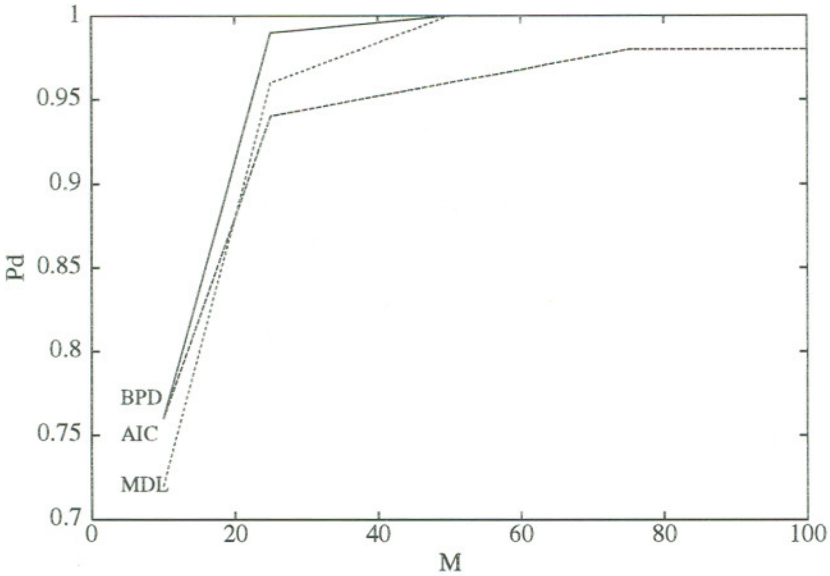

Fig. 1. Comparison of detection probabilities in term of the number of the snapshots. Two equal power (SNR = -3 dB) coherent signals with 90° phase difference, located at 15° and 20°, impinging on a linear array with six sensors (N=6).

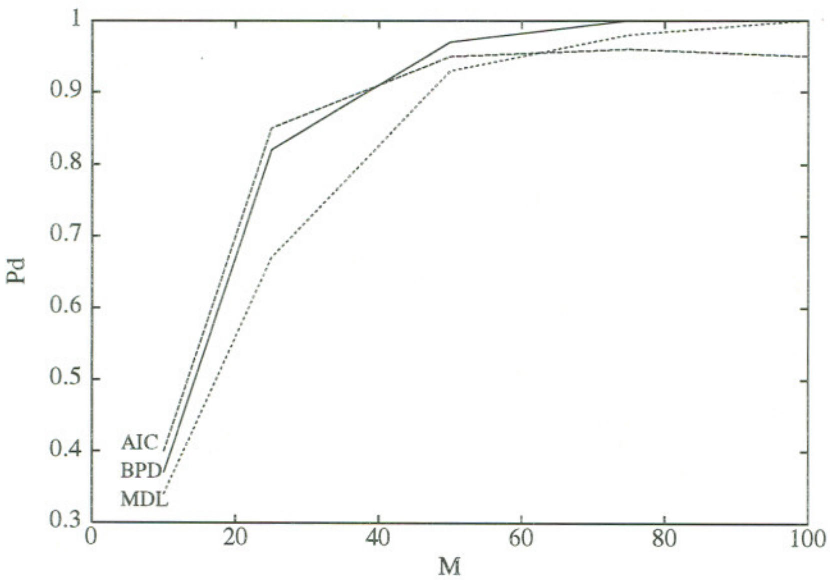

Fig. 2. Comparison of detection probabilities in term of the number of the snapshots. Two equal power (SNR = -3 dB) uncorrelated signals, located at 15° and 20°, impinging on a linear array with six sensors (N=6).

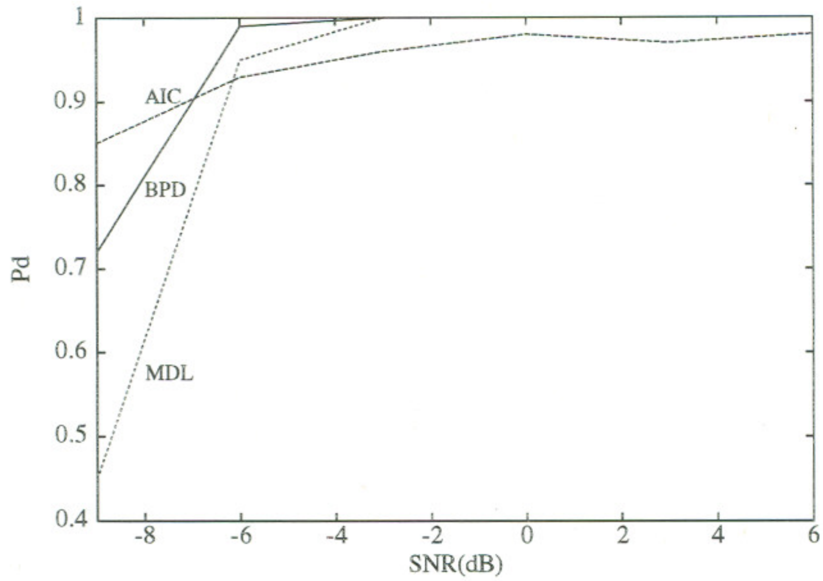

Fig. 3. Comparison of detection probabilities in term of the SNR. Two equal power coherent signals with 90° phase difference, located at 15° and 20°, impinging on a linear array with six sensors (N=6). The number of snapshots is 50.

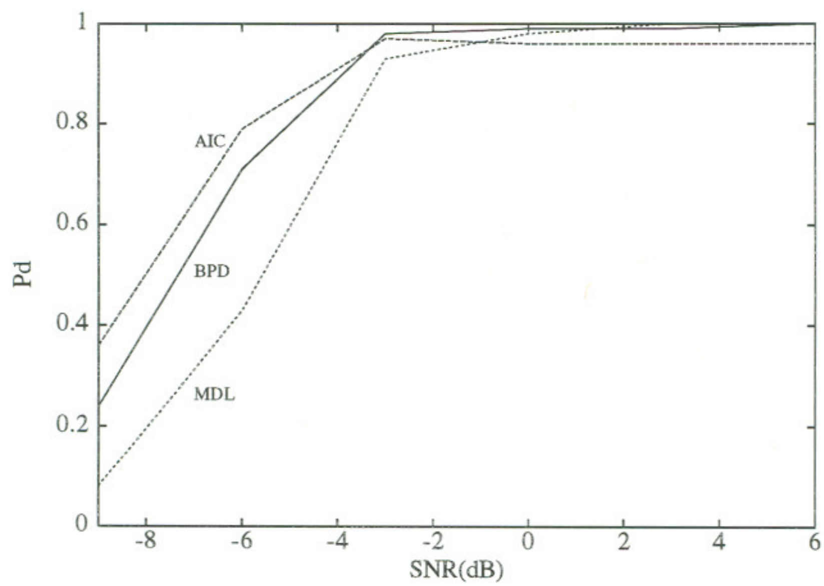

Fig. 4. Comparison of detection probabilities in term of the SNR. Two equal power uncorrelated signals, located at 15° and 20°, impinging on a linear array with six sensors (N=6). The number of snapshots is 50.

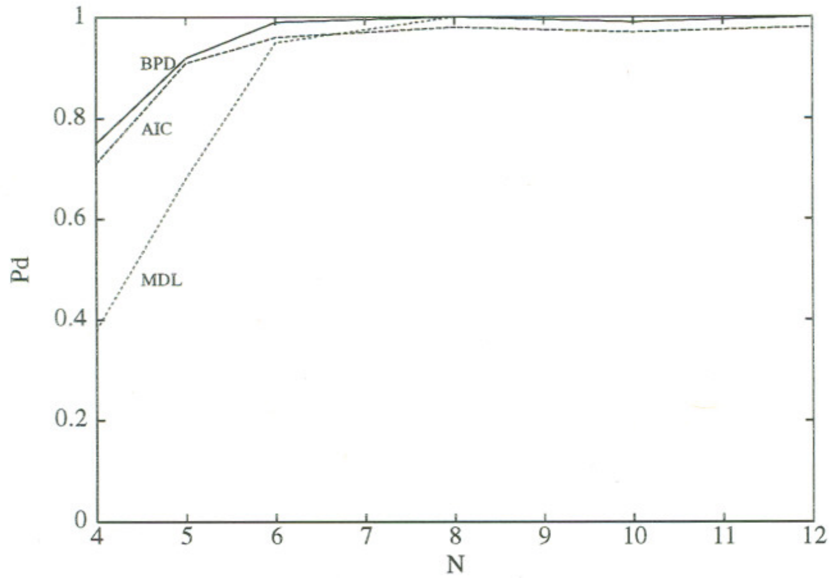

Fig. 5. Detection performance in term of the number of the sensors. Two equal power coherent signals with 90° phase difference, located at 10° and 20°, impinging on a linear array with N sensors. The SNR is -6 dB and the number of snapshots is 50.

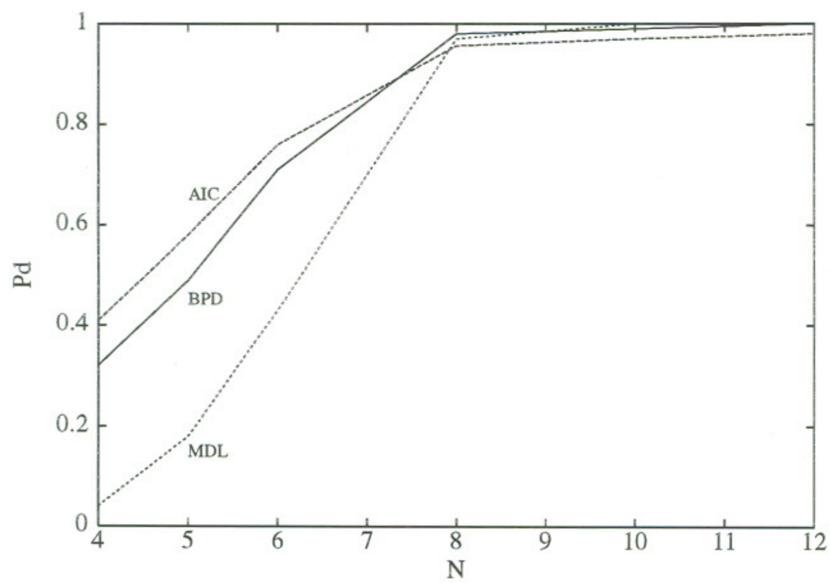

Fig. 6. Detection performance in term of the number of the sensors. Two equal power uncorrelated signals, located at 10° and 20°, impinging on a linear array with N sensors. The SNR is -6 dB and the number of snapshots is 50.

Appendix A:

In order to proof the consistency of the BPD criterion, we use the lemma:

Lemma A.1: Any information criteria given by

$$IC(k) = M \log C_{(M)}(\boldsymbol{\theta}_{(k)}) + \frac{1}{2}k(k+1)\alpha(M) \tag{A.1}$$

is strongly consistent if $\alpha(M) \rightarrow \infty$ and $\alpha(M)/M \rightarrow 0$ as $M \rightarrow \infty$.

Proof: see Zhao, Krishnaiah, and Bai [6], and Wax [7].

Ignoring the terms which are independent of M, the penalty function of the BPD criterion is expressed as

$$T'(k) = M(N-k)\ln(N-k) - \log\Gamma[M(N-k)] - \sum_{l=0}^{(k-1)} \log\Gamma[M-l] \tag{A.2}$$

In order to properly apply Lemma A.1, the penalty term has to satisfy the condition that the penalty is zero for $k = 0$. Since $T'(k) \neq 0$ as $k = 0$, we might obtain a relative penalty function such that $T''(k) = 0$ as $k = 0$:

$$\begin{aligned} T''(k) &= T'(k) - T'(0) \\ &= T_1 + T_2 + T_3 + T_4 + T_5 \end{aligned} \tag{A.3}$$

where

$$T_1 = M(N-k)\log(N-k) \tag{A.4}$$

$$T_2 = -\log\Gamma[\,M(N-k)\,] \tag{A.5}$$

$$\begin{aligned} T_3 &= -\sum_{l=0}^{k-1} \log\Gamma[\,M-l\,] \\ &\simeq -k\log\Gamma[\,M - \frac{k}{2}\,] \end{aligned} \tag{A.6}$$

for $M >> k$, and

$$T_4 = -MN\log N \tag{A.7}$$

$$T_5 = -\log\Gamma[\,MN\,] \tag{A.8}$$

For $M >> \log \pi$, we may apply the approximation:

$$\log \Gamma(x) \simeq (x - \frac{1}{2}) \log x \ - x \tag{A.9}$$

Therefore, (A.5), (A.6), and (A.8) can be rewitten as

$$\begin{aligned} T_2 &= -M(N-k)\log M(N-k) + \frac{1}{2}\log M(N-k) + M(N-k) \\ &= -M(N-k)\log(N-k) - MN\log M + Mk\log M + MN - Mk \\ &+ \frac{1}{2}\log M + \frac{1}{2}\log(N-k) \qquad \text{(A.10)} \\ T_3 &= -k((M - \frac{k}{2} - \frac{1}{2})\log(M - \frac{k}{2}) - (M - \frac{k}{2})) \\ &\simeq -KM\log M + \frac{1}{2}k(k+1)\log M + kM - k^2/2 \qquad \text{(A.11)} \end{aligned}$$

as $M >> k$ ($\log(M-k) \simeq \log M$), and

$$\begin{aligned} T_5 &= (MN - \frac{1}{2})\log MN - MN \\ &= MN\log M + MN\log N - \frac{1}{2}\log MN - MN \qquad \text{(A.12)} \end{aligned}$$

Summing up (A.4). (A.10), (A.11), (A.7), and (A.12), we get

$$T''(k) \doteq \frac{1}{2}k(k+1)\log M \ + \ \beta(N, L, k) \tag{A.13}$$

where $\beta(N, L, k)$ is the term which does not depend on M. Thus,

$$\alpha(M) \doteq \frac{M}{(M - N + 1)}\log M \ + \ \beta'(N, L, k) \tag{A.14}$$

and $\alpha(M) \to \infty$ and $\alpha(M)/M \to 0$ as $M \to \infty$. Therfore, according to Lemma A.1, the BPD criterion is shown to be strongly consistent. Q.E.D.

References

[1] H. Akaike, "A new look at the statistical model identification," *IEEE Trans. Automatic Control*, vol AC-19, no. 6, pp. 716-722, Dec. 1974.

[2] J. Rissanen, "A univeriversal prior for integers and estimation by minimum description length," *Annals of Statistics*, vol. 11, no. 2, pp. 416-431, 1983.

[3] G. Schwarz, "Estimation the dimension of a model," *Annals of Statistics*, vol. 6, no. 2, pp. 461-464, 1978.

[4] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-33, no. 2, pp. 387-392, Apr. 1985.

[5] K.M. Wong, Q. Zhang, J. P. Reilly, and P.C. Yip, "On onformation theoretic criteria for determining the number of signals in high resolution array processing," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 38, no. 11, pp. 1959-1971, Nov. 1990.

[6] L.C. Zhao, P.R. Krishnaiah, and Z.D. Bai, "On detection of the number of signals in presence of white noise," *Journal of Multivariate Analysis*, vol. 20, no. 1, pp. 1-25, Oct. 1986.

[7] M. Wax, "Detection and localization of multiple sources via the stochastic signals model," *IEEE Trans. Signal Processing*, vol. 39, no. 11, pp. 2450-2456, Nov. 1991.

[8] P.M. Djurić, *Selection of Signal and System Models by Bayesian Predictive Densities*, Ph.D. dissertation, Univ. of Rhode Island, 1990.

[9] J.F. Böhme, "Estimation of spectral parameters of correlated signals in wavefields," *Signal Processing*, vol. 11, pp. 329-337, 1986.

[10] A.G. Jaffer, "Maximum likelihood direction finding of stochastic sources: a separable solution," in Proc. ICASSP 88, pp. 2893-2296, 1988.

[11] R.L. Kashyap, "A Bayesian Comparison of different classes of dynamic models using empirical data," *Trans. Automatic Control*, vol AC-22, no. 5, Oct. 1977.

[12] G.E.P. Box and G.C. Tiao, *Bayesian Inference in Statistical Analysis*, Addison-Wesley, 1973.

[13] J. Lasenby and W.J. Fitzgerald, "A Bayesian approach to high-resolution beamforming," *IEE Proceedings-F*, vol. 138, no.6, pp. 539-544, Dec. 1991.

[14] A.P. Quinn, "The performance of Bayesian estimators in the superresolution of signal parameters," in Proc. ICASSP 92, vol. 5, pp.297-300, San Francisco, 1992.

[15] T. Leonard, "Comment on a paper by M. Lejeune and G.D. Faulkenberry," *Journal of American Statistical Association*, vol. 77, no. 379, pp. 657-658. Sep. 1982.

[16] W.Y. Tan, "On the complex analogue of Bayesian estimation of a multivariate regression model," *Annals of the Institute of Statistical Mathematics*, vol. 25, pp. 135-152, 1973.

[17] N.R. Goodman, "Statistical analysis based on a certain multivariate complex Gaussian distribution (an introduction)," *Ann. Math. Statist.*, vol. 34, pp. 152-177, 1963.

[18] B. Ottersten, M. Viberg, and T. Kailath, "Analysis of subspace fitting and ML techniques for parameter estimation from sensor array data," *IEEE Trans. Signal Processing*, vol. 40, no. 3, pp. 590-600, Mar. 1992.

[19] P. Stoica and A. Nehorai, "Performance study of conditional and unconditional direction-of-arrival estimation," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 38, no. 10, pp. 1783-1795, Oct. 1990.

[20] I. Ziskind and M. Wax, "Maximum likelihood localization of multiple sources by alternating projection," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 36, no. 10, pp. 1553-1560, Oct. 1988.

[21] S.D. Lin and M. Barkat, "Maximum likelihood estimation of direction of arrival and detection for broad band sources via dynamic programming," *IEEE Trans. Antennas and Propagation*, vol. 39, no. 8, pp. 1213-1221, Aug. 1991.

[22] D. Starer and A. Nehorai, "Newton algorithm for conditional and unconditional maximum likelihood estimation of the parameters of exponential signals in noise," *IEEE Trans. Signal Processing*, vol. 40, no. 6, pp. 1528-1534, June, 1992.