JOURNAL OF MULTIVARIATE ANALYSIS 20, 1–25 (1986)

On Detection of the Number of Signals in Presence of White Noise*

L. C. ZHAO, P. R. KRISHNAIAH, AND Z. D. BAI

Center for Multivariate Analysis, University of Pittsburgh

Communicated by the Editors

DEDICATED TO PROFESSOR C. R. RAO

In this paper, the authors propose procedures for detection of the number of signals in presence of Gaussian white noise under an additive model. This problem is related to the problem of finding the multiplicity of the smallest eigenvalue of the covariance matrix of the observation vector. The methods used in this paper fall within the framework of the model selection procedures using information theoretic criteria. The strong consistency of the estimates of the number of signals, under different situations, is established. Extensions of the results are also discussed when the noise is not necessarily Gaussian. Also, certain information-theoretic criteria are investigated for determination of the multiplicities of various eigenvalues.

1. INTRODUCTION

In the area of signal processing, it is of interest to detect the number of signals in presence of noise and estimate the parameters of the signals. The problem of estimation of the number of signals was discussed by Liggett [16], Schmidt [21], Tuft and Kumaresan [24], Wax, Shan, Kailath [27], and others in the literature. The model considered by them involves expressing the observation vector as the sum of Gaussian white noise and a vector of certain linear combinations of (random) signals radiated by sources. In this case, the number of signals is related to the

Received October 1, 1985.

AMS 1980 subject classifications: Primary 62H25, 62H15, 62H12.

Key words and phrases: complex elliptically symmetric distribution, consistency, detection of signals, eigenvalues, random matrices, signal detection and white noise.

* This work is supported by Contract N00014-85-K-0292 of the Office of Naval Research and Contract F49620-85-C-0008 of the Air Force Office of Scientific Research. The U.S. Government's right to retain a nonexclusive royalty-free license in and to the copyright covering this paper, for governmental purposes, is acknowledged.

0047-259X/86 $3.00

multiplicity of the smallest eigenvalue of the covariance matrix of the observation vector. The problem of testing the hypothesis of the multiplicity of the smallest eigenvalue of the covariance matrix was dealt extensively in multivariate statistical literature (e.g., see Anderson [2], Krishnaiah [10, 11] and Rao [17, 18]). Wax and Kailath [26] considered the problem of determination of the number of signals using information theoretic criteria proposed by Akaike [1], Rissanen [20], and Schwartz [22].

In the present paper, we use an alternative information theoretic criterion for detection of the number of signals and establish its consistency. In Section 2 of the paper, we state briefly the problems considered in this paper. In Sections 3 and 4, we establish the consistency of our procedures when the variance of the white noise is unknown and known, respectively, and the distribution underlying the observations is not necessarily complex Gaussian. Upper bounds on the probability of wrong detection of the procedures discussed in Sections 3 and 4 are given in a companion paper by Bai, Krishnaiah, and Zhao [4]. In Section 5, we discuss consistency of an alternative criterion for determination of the number of signals. Some remarks are also made for the case when the observations are distributed as complex elliptically symmetric.

Here, we note that complex elliptically symmetric distribution was introduced by Krishnaiah and Lin [12]. In Section 6, we propose an information theoretic criterion for determination of the number of signals and the multiplicities of the eigenvalues of the covariance matrix which are different from the smallest eigenvalue when the variance of white noise is unknown. The consistency of the above procedure is also established. In Section 7, results analogous to those proved in Section 6 are established when the variance of white noise is known. The problem of detection of the number of signals when the noise covariance matrix is arbitrary is discussed by Zhao, Krishnaiah, and Bai [31] in a companion paper when an independent estimate of the above covariance matrix is available.

2. Preliminaries and Statement of Problems

Consider the model

$$\mathbf{x}(t) = A\mathbf{s}(t) + \mathbf{n}(t) \tag{2.1}$$

where $A = [A(\mathbf{\Phi}_1), ..., A(\mathbf{\Phi}_q)]$, $\mathbf{s}(t) = (s_1(t), ..., s_q(t))'$, $\mathbf{n}(t) = (n_1(t), ..., n_p(t))'$ and $q < p$. In the above model, $\mathbf{n}(t)$ is the noise vector distributed independent of $\mathbf{s}(t)$ as complex multivariate normal with mean vector $\mathbf{0}$ and covariance matrix $\sigma^2 I_p$. Also, $\mathbf{s}(t)$ is distributed as complex multivariate normal with mean vector $\mathbf{0}$ and nonsingular covariance matrix Ψ and $A(\mathbf{\Phi}_i): p \times 1$ is a complex vector of functions of the elements of unknown vector $\mathbf{\Phi}_i$ associated with ith signal. Also, $s_i(t)$ is the waveform associated with ith signal. Then, the covariance matrix Σ of $\mathbf{x}(t)$ is given by

$$\Sigma = A\Psi\bar{A}' + \sigma^2 I_p \tag{2.2}$$

where $\bar{A}'$ denotes the transpose of the complex conjugate of A. We assume that $\mathbf{x}(t_1),..., \mathbf{x}(t_N)$ are independent observations on $\mathbf{x}(t)$ unless stated otherwise. Now, let $\lambda_1 \geqslant \cdots \geqslant \lambda_p$ denote the eigenvalues of Σ, and $\theta_1 \geqslant \cdots \geqslant \theta_q$ denote the nonzero eigenvalues of $A\Psi\bar{A}'$. Also, let H_q denote the hypothesis $\lambda_q > \lambda_{q+1} = \cdots = \lambda_p = \sigma^2$. Under H_q, $\lambda_i = \sigma^2 + \theta_i$ $(i = 1, 2,..., q)$ and $\lambda_{q+j} = \sigma^2$ $(j = 1, 2,..., p-q)$. So, H_q is equivalent to the hypothesis that q signals are transmitted. Wax and Kailath [26] used Akaike's AIC criterion and Schwart–Rissanen's minimum description length (MDL) criterion for model selection for determination of the value of q. According to the AIC criterion, q is estimated with $\hat{q}$ where $\hat{q}$ is chosen such that

$$\mathrm{AIC}(\hat{q}) = \min\{\mathrm{AIC}(0),..., \mathrm{AIC}(p-1)\} \tag{2.3}$$

and

$$\mathrm{AIC}(k) = -2\log L_k + 2v(k, p). \tag{2.4}$$

Here L_k is the likelihood ratio test statistic for testing H_k against the alternative that Σ is arbitrary, and $v(k, p)$ denotes the number of free parameters that have to be estimated under H_k. According to the MDL criterion, the value of q is estimated with $\hat{q}$ where $\hat{q}$ is chosen such that

$$\mathrm{MDL}(\hat{q}) = \min\{\mathrm{MDL}(0),..., \mathrm{MDL}(p-1)\} \tag{2.5}$$

$$\mathrm{MDL}(k) = -\log L_k + \frac{\log N}{2} v(k, p). \tag{2.6}$$

In the present paper, we consider the following alternative information theoretic criterion for model selection for estimation of the value of q. According to this new information theoretic criterion for model selection, we estimate q with $\hat{q}$ where $\hat{q}$ is chosen such that

$$I(\hat{q}, C_N) = \min\{I(0, C_N),..., I(p-1, C_N)\} \tag{2.7}$$

$$I(k, C_N) = -\log L_k + C_N v(k, p) \tag{2.8}$$

and C_N is chosen such that

$$\lim_{N\to\infty} \{C_N/N\} = 0 \tag{2.9}$$

$$\lim_{N\to\infty} \{C_N/\log\log N\} = \infty. \tag{2.10}$$

If the assumption of normality and independence of observations $x(t_1),..., x(t_N)$ are violated, L_k is no longer the likelihood ratio test statistic. But, we can still use the criterion defined by (2.7)–(2.10) when L_k denotes the likelihood ratio test statistic for testing H_k under the assumptions of normality and independence of observations. The probability of correct detection of the procedure proposed by us is given by

$$P(CD) = P[I(q, C_N) - I(k, C_N) < 0; \qquad k = 0, 1,..., (p-1); k \neq q \mid H_q].$$

We are interested in establishing the strong consistency of the above procedure for the cases when σ^2 is unknown and known under certain assumptions about the underlying distribution.

3. Consistency of $I(\hat{q}, C_N)$ Criterion When σ^2 Is Unknown

In this section, we establish the consistency of the estimate $\hat{q}$ of q when the criterion $I(\hat{q}, C_N)$ is used and σ^2 is unknown when the underlying distribution is not necessarily complex multivariate normal. The main result of this section is stated in the following theorem:

THEOREM 3.1. *Suppose* $\mathbf{x}(t)$ *is a complex, stationary process with* $E(\mathbf{x}(t)) = 0$ *and* $E(\bar{\mathbf{x}}'(t)\,\mathbf{x}(t))^2 < \infty$. *Also, we assume that* $\{\mathbf{x}(t_i),\ i = 1, 2,...,\}$ *is a stationary and* ϕ*-mixing sample sequence with* ϕ *decreasing and* $\sum_{j=1}^{\infty} \phi^{1/2}(j) < \infty$. *Let* $\hat{q}$ *be chosen such that*

$$I(\hat{q}, C_N) = \min\{I(0, C_N),..., I(p-1, C_N)\} \tag{3.1}$$

where $I(k, C_N)$ *was defined by* (2.8) *and* C_N *is chosen satisfying* (2.9) *and* (2.10). *Then* $\hat{q}$ *is a strongly consistent estimate of* q.

We need the following results to prove the above theorem.

LEMMA 3.1. *Suppose* $\{x_i, i \geqslant 1\}$ *is a stationary real* ϕ*-mixing sequence* *with* $E(x_1)=0$ *and* $E(|x_1|^2)<\infty$. *Also,* ϕ *is decreasing with* $\sum_{j=1}^{\infty}\phi^{1/2}(j)<\infty$. *Then*

$$\limsup_{n\to\infty}\left\{\sum_{i=1}^{n} x_i/(2n\delta^2\log\log n\delta^2)^{1/2}\right\}=1 \qquad \text{a.s.} \tag{3.2}$$

where $\delta^2=Ex_1^2+2\sum_{i=1}^{\infty}Ex_1x_{1+i}\neq 0$.

We note that $\sum_{j=1}^{\infty}\phi^{1/2}(j)<\infty$ implies $\delta^2<\infty$. A proof of the above lemma is given in Hall and Heyde [7, p. 145]. For some earlier work on this topic, the reader is referred to Reznik [19] and Stout [23].

Remark. If $\delta^2=0$, then

$$\lim_{n\to\infty}\left\{\sum_{i=1}^{n} x_i/\sqrt{2n\log\log n}\right\}=0 \qquad \text{a.s.} \tag{3.2$'$}$$

To see this, let $\{y_n\}$ be a sequence of i.i.d. random variables independent of $\{x_n\}$ such that $E(y_1)=0$ and $\tilde{\delta}^2=E(y_1^2)>0$. Then (3.2) holds for $\{y_n\}$ and $\{x_n+y_n\}$, where δ^2 in (3.2) is replaced with $\tilde{\delta}^2$. Since $\tilde{\delta}^2$ can be arbitrarily small, (3.2$'$) follows.

LEMMA 3.2. *Suppose that* A, A_n, $n=1,2,\ldots$, *are all* $p\times p$ *symmetric matrices such that* $A_n-A=O(\alpha_n)$ *and* $\alpha_n\to 0$ *as* $n\to\infty$. *Denote by* $\lambda_1\geqslant\lambda_2\geqslant\cdots\geqslant\lambda_p$ *and* $\lambda_1^{(n)}\geqslant\cdots\geqslant\lambda_p^{(n)}$ *the eigenvalues of* A *and* A_n, *respectively. Then we have*

$$\lambda_i^{(n)}-\lambda_i=O(\alpha_n) \qquad \text{as} \quad n\to\infty,\ i=1,\ldots,p.$$

Proof. Without loss of generality, we can assume that $A=\operatorname{diag}[\tilde{\lambda}_1 I_{\mu_1},\ldots,\tilde{\lambda}_r I_{\mu_r}]$, where $\tilde{\lambda}_1>\tilde{\lambda}_2>\cdots>\tilde{\lambda}_r$. According to Bai [3], we know $\lambda_i^{(n)}-\lambda_i\to 0$. First we consider the special case where $r=1$. For any i,

$$\begin{aligned} 0=|\lambda_i^{(n)}I_p-A_n| &= |(\lambda_i^{(n)}-\tilde{\lambda}_1)I_p-(A_n-A)| \\ &= (\lambda_i^{(n)}-\tilde{\lambda}_1)^p+\sum_{l=1}^{p}(-1)^l(\lambda_i^{(n)}-\tilde{\lambda}_1)^{p-l}D_l, \end{aligned} \tag{3.3}$$

where D_l is the sum of all l-ordered principal minors of A_n-A. Since $A_n-A=O(\alpha_n)$, we have $D_l=O(\alpha_n^l)$. By (3.3) we know

$$|\lambda_i^{(n)}-\tilde{\lambda}_1|\leqslant\sum_{l=1}^{p}|\lambda_i^{(n)}-\tilde{\lambda}_1|^{-l+1}O(\alpha_n^l)$$

which implies $|\lambda_i^{(n)}-\tilde{\lambda}_1|=O(\alpha_n)$ as $n\to\infty$ for $i=1,\ldots,p$.

Now we consider the general case. Suppose $i \leqslant \mu_1$. We have

$$
\begin{aligned}
0 &= |\lambda_i^{(n)} I_p - A_n| \\
&= \left| \begin{pmatrix} (\lambda_i^{(n)} - \tilde{\lambda}_1) I_{\mu_1} & 0 & \cdots & 0 \\ 0 & (\lambda_i^{(n)} - \tilde{\lambda}_2) I_{\mu_2} & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \cdots & (\lambda_i^{(n)} - \tilde{\lambda}_r) I_{\mu_r} \end{pmatrix} - (A_n - A) \right| \\
&\triangleq \begin{vmatrix} (\lambda_i^{(n)} - \tilde{\lambda}_1) I_{\mu_1} - B_{11}^{(n)} & -B_{12}^{(n)} \\ -B_{21}^{(n)} & B_{22}^{(n)} \end{vmatrix} \\
&= |B_{22}^{(n)}| \cdot |(\lambda_i^{(n)} - \tilde{\lambda}_1) I_{\mu_1} - B_{11}^{(n)} - B_{12}^{(n)} B_{22}^{(n)-1} B_{21}^{(n)}|.
\end{aligned}
$$

Since $B_{22}^{(n)} \to \operatorname{diag}[(\tilde{\lambda}_1 - \tilde{\lambda}_2) I_{\mu_2}, ..., (\tilde{\lambda}_1 - \tilde{\lambda}_r) I_{\mu_r}]$, $B_{22}^{(n)}$ is nonsingular for all large n. Thus

$$|(\lambda_i^{(n)} - \tilde{\lambda}_1) I_{\mu_1} - B_{11}^{(n)} - B_{12}^{(n)} B_{22}^{(n)-1} B_{21}^{(n)}| = 0. \tag{3.4}$$

From $A_n - A = O(\alpha_n)$, it follows that $B_{11}^{(n)} = O(\alpha_n)$ and $B_{12}^{(n)} B_{22}^{(n)-1} B_{21}^{(n)} = O(\alpha_n^2)$. Using the result proved just before, we get

$$\lambda_i^{(n)} - \tilde{\lambda}_1 = O(\alpha_n)$$

for $i = 1, ..., \mu_1$. By the same approach, we can prove

$$\lambda_i^{(n)} - \tilde{\lambda}_h = O(\alpha_n), \quad i = \mu_1 + \cdots + \mu_{h-1} + 1, ..., \mu_1 + \cdots + \mu_h, \ h = 1, ..., r, \tag{3.5}$$

which complete the proof of the lemma.

Let $\delta_1 \geqslant \cdots \geqslant \delta_p$ denote the eigenvalues of $\hat{\Sigma}$, where $N\hat{\Sigma} = \sum_{i=1}^{N} \mathbf{x}(t_i) \bar{\mathbf{x}}'(t_i)$. Using Lemma 3.1 and the conditions imposed on $\{\mathbf{x}(t_i), i = 1, 2, ..., \}$, we have

$$\hat{\Sigma} - \Sigma = O(\sqrt{\log \log N/N}) \qquad \text{a.s.} \tag{3.6}$$

Now, applying Lemma 3.2, we obtain

$$\delta_j - \lambda_j = O(\sqrt{\log \log N/N}) \qquad \text{a.s.} \tag{3.7}$$

for $j = 1, 2, ..., p$.

When $\mathbf{x}(t)$ is distributed as complex multivariate normal and the observations are independent, the likelihood function for testing the hypothesis H_k against the alternative that Σ has general structure is known to be

$$L_k = \left\{ \prod_{i=k+1}^{p} \delta_i^N \Big/ \left(\frac{1}{p-k} \sum_{i=k+1}^{p} \delta_i \right)^{N(p-k)} \right\}. \tag{3.8}$$

We will prove the consistency of the method based upon the criterion $I(\hat{q}, C_N)$. Let $G_1(k) = \log L_k$ and

$$G(k) = G_1(k) - C_N[k(2p-k)+1] \tag{3.9}$$

where $k(2p-k)+1$ is the number of free parameters that have to be estimated under the hypothesis H_k and L_k is given by (3.8). Assume $k < q$. Using (3.7), we get

$$\lim_{N\to\infty} \frac{1}{N}(G_1(q) - G_1(k)) = W(q, k) \qquad \text{a.s.} \tag{3.10}$$

where

$$\begin{aligned} W(q, k) &= \log\left(\prod_{i=q+1}^{p} \lambda_i\right) - (p-q)\log\left(\frac{1}{(p-q)}\sum_{i=q+1}^{p} \lambda_i\right) \\ &\quad - \log\left(\prod_{i=k+1}^{p} \lambda_i\right) + (p-k)\log\left(\frac{1}{(p-k)}\prod_{i=k+1}^{p} \lambda_i\right). \qquad (3.11) \\ &= (q-k)\log\left[\frac{1}{(q-k)}\sum_{i=k+1}^{q} \lambda_i \Big/ \left(\prod_{i=k+1}^{q} \lambda_i\right)^{1/(q-k)}\right] \\ &\quad + (p-k)[\log(\alpha_1 A_1 + \alpha_2 A_2) - (\alpha_1 \log A_1 + \alpha_2 \log A_2)] \end{aligned}$$

where

$$\alpha_1 = (q-k)/(p-k), \qquad \alpha_2 = (p-q)/(p-k)$$

$$A_1 = \frac{1}{(q-k)}\sum_{i=k+1}^{q} \lambda_i, \qquad A_2 = \frac{1}{(p-q)}\sum_{i=q+1}^{p} \lambda_i.$$

By the well-known arithmetic mean geometric mean inequatity, we have

$$W(q, k) \geqslant (p-k)[\log(\alpha_1 A_1 + \alpha_2 A_2) - (\alpha_1 \log A_1 + \alpha_2 \log A_2)]. \tag{3.12}$$

Also, $A_1 > A_2$. By Jensen's inequality, we have

$$W(q, k) > 0. \tag{3.13}$$

Using (3.9), (3.10), (3.13), and $\lim_{N\to\infty} (C_N/N) = 0$, we obtain

$$G(q) - G(k) = NW(q, k)(1 + o(1)) \qquad \text{a.s.}$$

So, with probability one for large N, we have

$$G(q) > G(k). \tag{3.14}$$

Now we assume $k > q$ and $k \leqslant p-1$. Without loss of generality we can assume $\sigma^2 = 1$. By (3.7) we have $\lim_{N \to \infty} (\delta_j - 1) = 0$ a.s. for $j = q+1, ..., p$. Using Taylor's expansion, we get for $k > q$

$$\begin{aligned} G_1(k) &= N \left\{ \sum_{i=k+1}^{p} \log(1 + \delta_i - 1) - (p-k) \log\left(1 + \frac{1}{p-k} \sum_{i=k+1}^{p} (\delta_i - 1)\right)\right\} \\ &= -\frac{N}{2} \sum_{i=k+1}^{p} (\delta_i - 1)^2 (1 + o(1)) \\ &\quad + \frac{N}{2(p-k)} \left(\sum_{i=k+1}^{p} (\delta_i - 1)\right)^2 (1 + o(1)) \qquad \text{a.s.} \end{aligned}$$

By (3.7) we see that

$$\begin{aligned} G_1(k) &= O(\log\log N) \qquad \text{a.s.,} \quad p-1 \geqslant k > q \\ G_1(q) &= O(\log\log N) \qquad \text{a.s.} \end{aligned} \tag{3.15}$$

From (3.9), (3.15), and $C_N/\log\log N \to \infty$, we get

$$\begin{aligned} G(q) - G(k) &= C_N(k-q)(2p-k-q) + O(\log\log N) \\ &= C_N(k-q)(2p-k-q)(1+o(1)) \qquad \text{a.s.} \end{aligned} \tag{3.16}$$

Thus with probability one for large N we have

$$G(q) > G(k). \tag{3.17}$$

From (3.14) and (3.17), it follows that with probability one for large N

$$\hat{q} = q.$$

Thus the proof of Theorem 1 is completed.

When $\mathbf{x}(t)$ is distributed as real multivariate normal, the proof goes along the same lines as in the complex case. In general, Theorem 3.1 is true for the real stationary process.

Wax and Kailath [26] stated that the MDL criterion is strongly consistent but the AIC criterion is not. To establish the above statements, they claimed that for $k > q$, $-2\{\log L_q - \log L_k\}$ is asymptotically chi-square with $(k-q)(2p-k-q)$ degrees of freedom. But, Zhao, Krishnaiah, and Bai [30] showed that the above claim is incorrect. However, the consistency of the MDL criterion follows from our results since the MDL criterion is a special case of our criterion.

4. Detection of the Number of Signals When Variance of White Noise Is Known

In Section 3, we discussed a model selection criterion for detection of the number of signals when the distribution underlying the observations is complex multivariate normal and the variance of white noise is unknown. In this section, we derive analogous criterion when the underlying distribution is (real) multivariate normal and the variance of the white noise is known. The strong consistency of the above criterion is also established.

In the model (2.1), we assume that the noise vector $\mathbf{n}(t)$ is distributed as the multivariate normal with mean vector $\mathbf{0}$ and covariance matrix $\sigma^2 I_p$, A is a real matrix of rank $q < N$, and the signal vector $\mathbf{s}(t)$ is distributed independent of $\mathbf{n}(t)$ as a multivariate normal with mean vector $\mathbf{0}$ and nonsingular covariance matrix Ψ. Then, the covariance matrix of $\mathbf{x}(t)$ is $\Sigma = A\Psi A' + \sigma^2 I$. We assume that σ^2 is known. Without loss of generality, we assume that $\sigma^2 = 1$. Let $\lambda_1 \geqslant \cdots \geqslant \lambda_p$ denote the eigenvalues of Σ. Now, let

$$H_k^*: \lambda_1 \geqslant \cdots \geqslant \lambda_k > \lambda_{k+1} = \cdots = \lambda_p = 1. \tag{4.1}$$

The kth model M_k^* is the one for which H_k^* is true. We are interested in selecting one of the p models $M_0^*, M_1^*, ..., M_{p-1}^*$.

When the observations $\mathbf{x}(t_1), ..., \mathbf{x}(t_N)$ are independent, the logarithm of the likelihood function is given by

$$\log L(\theta) = -\frac{N}{2} \log |\Sigma| - \frac{N}{2} \operatorname{tr} \Sigma^{-1}\hat{\Sigma} \tag{4.2}$$

where

$$\hat{\Sigma} = \sum_{j=1}^{N} \mathbf{x}_j \mathbf{x}_j'/N, \quad \mathbf{x}_j = \mathbf{x}(t_j). \tag{4.3}$$

Also, let $\delta_1 \geqslant \cdots \geqslant \delta_p$ be the eigenvalues of $\hat{\Sigma}$. In addition, let τ denote the number of δ_i's which are greater than one. Also, let $d \leqslant \tau$. We will first calculate $L^*(\lambda_{d+1}, ..., \lambda_p) = \sup_{\phi_d} \log L(\theta)$, where $\sup_{\phi_d} \log L(\theta)$ indicates that $\log L(\theta)$ is maximized subject to the condition that $\lambda_1 \geqslant \cdots \geqslant \lambda_d > 1$.

Write $\Lambda = \operatorname{diag}(\lambda_1, ..., \lambda_p)$, $\Delta = \operatorname{diag}(\delta_1, ..., \delta_p)$. There exist two real orthogonal matrices $\mathcal{O}_1$ and $\mathcal{O}_2$ such that

$$\Sigma = \mathcal{O}_1' \Lambda \mathcal{O}_1, \qquad \hat{\Sigma} = \mathcal{O}_2' \Delta \mathcal{O}_2.$$

Put $Q = \mathcal{O}_2 \mathcal{O}_1'$. Then we have

$$\log L(\theta) = -\frac{N}{2} \sum_{j=1}^{p} \log \lambda_j - \frac{N}{2} \operatorname{tr} \Lambda^{-1} Q' \Delta Q.$$

Since Q is orthogonal, we have

$$\operatorname{tr} \Lambda^{-1} Q' \Delta Q \geqslant \sum_{j=1}^{p} \delta_i/\lambda_i,$$

and the equality holds for $Q = I_p$ (see Von Neumann [25]). So,

$$\sup_{\Phi_d} \log L(\theta) = \sup_{\Phi_d} \left\{ -\frac{N}{2} \sum_{j=1}^{p} \log \lambda_j - \frac{N}{2} \sum_{j=1}^{p} (\delta_j/\lambda_j) \right\} \tag{4.4}$$

i.e.,

$$\begin{aligned} L^*(\lambda_{d+1}, ..., \lambda_p) &= -\frac{N}{2} \log(\lambda_{d+1} \cdots \lambda_p) - \frac{N}{2} \sum_{j=d+1}^{p} \delta_j/\lambda_j \\ &\quad + \sup_{\Phi_d} \left\{ -\frac{N}{2} \log (\lambda_1 \cdots \lambda_d) - \frac{N}{2} \sum_{j=1}^{d} \delta_j/\lambda_j \right\} \\ &= -\frac{N}{2} \log (\lambda_{d+1} \cdots \lambda_p) - \frac{N}{2} \sum_{j=d+1}^{p} \delta_j/\lambda_j \\ &\quad - \frac{N}{2} \log (\delta_1 \cdots \delta_d) - \frac{N}{2} d \end{aligned} \tag{4.5}$$

where the supremum is attained at $\delta_j = \lambda_j$ for $j = 1, 2, ..., d$.

First, we assume that $\tau < k$. In this case

$$\sup_{\theta \in \Theta_k} \log L(\theta) = \sup_{\Phi(\tau,k)} L^*(\lambda_{\tau+1}, ..., \lambda_p) \tag{4.6}$$

where Θ_k denotes the parametric space when H_k^* is true, $\sup_{\Phi(\tau,k)}$ indicates that the supremum is taken over $\lambda_{\tau+1} \geqslant \cdots \geqslant \lambda_k > 1$ and $\lambda_{k+1} = \cdots = \lambda_p = 1$. But

$$\begin{aligned} &\sup_{\Phi(\tau,k)} L^*(\lambda_{\tau+1}, ..., \lambda_p) \\ &\quad = -\frac{N}{2} \sum_{i=1}^{\tau} \log \delta_i - \frac{N}{2} \tau - \frac{N}{2} \sum_{i=k+1}^{p} \delta_i + \frac{N}{2} \sup_{\lambda_{\tau+1} \geqslant \cdots \geqslant \lambda_k > 1} \\ &\quad\quad \times \left\{ - \sum_{i=\tau+1}^{k} \log \lambda_i - \sum_{j=\tau+1}^{k} \delta_i/\lambda_i \right\}. \end{aligned} \tag{4.7}$$

Also, $\delta_i \leqslant 1$ and $\lambda_i > 1$ for $i = \tau+1, ..., k$. So,

$$-(\log \lambda_i + (\delta_i/\lambda_i)) < -\delta_i \tag{4.8}$$

and the equality holds only when $\lambda_i = 1$. Since the above λ_i's can be arbitrarily approximated to one, we have

$$\sup_{\theta \in \Theta_k} \log L(\theta) = -\frac{N}{2} \sum_{i=1}^{\tau} \log \delta_i - N\tau/2 - \frac{N}{2} \sum_{i=\tau+1}^{p} \delta_i \tag{4.9}$$

when $\tau < k$. Next, let $\tau \geqslant k$. Then

$$\begin{aligned} \sup_{\theta \in \Theta_k} \log L(\theta) &= -\frac{N}{2} \log(\delta_1 \cdots \delta_k) - \frac{N}{2} k \\ &\quad + \sup_{\lambda_{k+1} = \cdots = \lambda_p = 1} \left\{ -\frac{N}{2} \log(\lambda_{k+1} \cdots \lambda_p) - \frac{N}{2} \sum_{j=k+1}^{p} \delta_j/\lambda_j \right\} \\ &= -\frac{N}{2} \sum_{i=1}^{k} \log \delta_i - \frac{N}{2} k - \frac{N}{2} \sum_{i=k+1}^{p} \delta_i. \end{aligned} \tag{4.10}$$

Combining (4.9) and (4.10), we obtain

$$\sup_{\theta \in \Theta_k} \log L(\theta) = -\frac{N}{2} \sum_{i=1}^{p} \log \delta_i - \frac{Np}{2} + \frac{N}{2} \sum_{i=1+\min(\tau,k)}^{p} (\log \delta_i + 1 - \delta_i). \tag{4.11}$$

But the supremum of $L(\theta)$ over the whole parametric space is given by

$$-\frac{N}{2} \sum_{i=1}^{p} \log \delta_i - \frac{Np}{2}.$$

So, the logarithm of the likelihood ratio test statistic for testing H_k^* is given by

$$\log L_k = \frac{N}{2} \sum_{i=1+\min(\tau,k)}^{p} (\log \delta_i + 1 - \delta_i). \tag{4.12}$$

Now, let

$$\log \tilde{L}_k = \frac{N}{2} \sum_{i=k+1}^{p} (\log \delta_i + 1 - \delta_i). \tag{4.13}$$

We know from (3.7) that

$$\delta_i - \lambda_i = O\left(\left(\frac{\log \log N}{N}\right)^{1/2}\right) \quad \text{a.s.} \tag{4.14}$$

Suppose the true model is M_q. Then

$$\lambda_1 \geqslant \cdots \geqslant \lambda_q > \lambda_{q+1} = \cdots = \lambda_p = 1. \tag{4.15}$$

From (4.14), we know with probability one, that $\delta_i > 1$ for $i = 1, 2, \ldots, q$ and $\min(q, \tau) = q$ for large N. So, the statistics $\log L_q$ and $\log \tilde{L}_q$ have the same distribution asymptotically. Here, we note that Anderson [2] suggested to use $\tilde{L}_q$ as a statistic to test H_q^* and pointed out that the asymptotic distribution of $-2 \log \tilde{L}_q$ is chi-square with $(p-q)(p-q+1)/2$ degrees of freedom. Rao [18] pointed out that $\tilde{L}_q$ is not the LRT statistic.

We will now consider the problem of selecting one of the models $M_0, M_1, \ldots, M_{p-1}$ by using an information theoretic criterion. Let

$$G(k) = \log L_k - C_N k(2p - k + 1)/2 \tag{4.16}$$

where C_N satisfies the following conditions

(i) $\lim_{N \to \infty} (C_N/N) = 0$,

(ii) $\lim_{N \to \infty} (C_N/\log \log N) = \infty$.

We select the model $M_{\hat{q}}$ where $\hat{q}$ is chosen such that

$$G(\hat{q}) = \max_{0 \leqslant k \leqslant p-1} G(k). \tag{4.17}$$

We will now show that $\hat{q}$ is a consistent estimate of q.

THEOREM 4.1. *If $N\hat{\Sigma}$ is distributed as central Wishart matrix with N degrees of freedom and $E(\hat{\Sigma}) = \Sigma$, then $\hat{q}$ is a strongly consistent estimate of q.*

Proof. Suppose that M_q is the true model and $k < q$. We have

$$G(q) - G(k) = \log L_q - \log L_k - C_N(q-k)(2p-k-q+1)/2. \tag{4.18}$$

As mentioned above, with probability one, we have for large N,

$$\delta_i > 1, \ i = 1, \ldots, q \qquad \text{and} \qquad \min(q, \tau) = q. \tag{4.19}$$

Thus with probability one for large N,

$$\begin{aligned} \log L_q - \log L_k &= \tfrac{1}{2}N \sum_{i=q+1}^{p} (\log \delta_1 + 1 - \delta_i) - \tfrac{1}{2}N \sum_{i=k+1}^{p} (\log \delta_i + 1 - \delta_i) \\ &= -\tfrac{1}{2}N \sum_{i=k+1}^{q} (\log \delta_i + 1 - \delta_i) = \tfrac{1}{2}N W_N(q, k), \end{aligned}$$

where

$$W_N(q, k) = -\sum_{i=k+1}^{q} (\log \delta_i + 1 - \delta_i).$$

We have

$$\lim_{N\to\infty} W_N(q,k) \overset{\text{a.s.}}{=} W(q,k) \equiv -\sum_{i=k+1}^{q} (\log\lambda_i + 1 - \lambda_i) > 0.$$

Hence, with probability one, we have for large N,

$$\log L_q - \log L_k > \tfrac{1}{4} N W(q,k),$$

and

$$G(q) - G(\mathrm{k}) > 0. \tag{4.20}$$

Here we used the condition $\lim_{N\to\infty} C_N/N = 0$.

Now we assume that $k > q$. By (4.19) we have

$$|\log L_q - \log L_k| \leqslant N \sum_{i=q+1}^{p} |\log\delta_i + 1 - \delta_i|.$$

Since $|\delta_i - 1| = O((\log\log N/N)^{1/2})$ a.s. for $i > q$, we can use Taylor's expansion, to get

$$\begin{aligned} |\log L_q - \log L_k| &\leqslant N \sum_{i=q+1}^{p} \tfrac{1}{2}(\delta_i - 1)^2 (1 + o(1)) \qquad \text{a.s.} \\ &= O(\log\log N) \qquad \text{a.s.} \end{aligned}$$

From $(C_N/\log\log N) \to \infty$, we see that with probability one, for large N,

$$G(q) - G(k) = O(\log\log N) + C_N(k-q)(2p-k-q+1)/2 > 0. \tag{4.21}$$

From (4.20) and (4.21), it follows that with probability one for large N,

$$\hat{q} = q.$$

Thus Theorem 4.1 is proved.

When the underlying distribution is complex multivariate normal, the proof for the consistency of the method goes along the same lines as in the real case.

5. Further Results on Determination of the Number of Signals

We will first discuss procedures for determination of the number of signals transmitted when the underlying distribution is real or complex elliptically symmetric. Here, we note that a random vector $\mathbf{y}$ is said to be elliptically symmetric if its density is of the form

$$f(\mathbf{y}) = |\Sigma|^{-1/2} g(\tfrac{1}{2}(\mathbf{y}-\boldsymbol{\mu})' \Sigma^{-1}(\mathbf{y}-\boldsymbol{\mu})) \tag{5.1}$$

where g is a non-increasing function in $[o, \infty)$. Multivariate normal and multivariate t distributions are special cases of the elliptically symmetric distributions. Kelker [9] proposed the elliptically symmetric distributions and studied some of its properties. Krishnaiah and Lin [12] proposed complex elliptically symmetric distribution and studied some of its properties. A complex random vector $\mathbf{x} = \mathbf{x}_1 + i\mathbf{x}_2$ is said to be distributed as complex elliptically symmetric distribution if its density is of the form

$$f(\mathbf{x}) = |\Sigma|^{-1} h((\overline{\mathbf{x} - \boldsymbol{\mu}})' \Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu})) \tag{5.2}$$

where Σ is Hermitian, $\bar{\mathbf{a}}$ denotes the complex conjugate of $\mathbf{a}$, and $h(\cdot)$ is a non-increasing function in $[\mathbf{o}, \infty)$. The covariance matrix of $(\mathbf{x}_1', \mathbf{x}_2')$ has the structure

$$\begin{pmatrix} \Sigma_1 & \Sigma_2 \\ -\Sigma_2 & \Sigma_1 \end{pmatrix}.$$

Complex multivariate normal considered by Wooding [29] and Goodman [6] and complex multivariate t distribution are special cases of the complex elliptically symmetric distribution. The density of the complex multivariate normal is known to be

$$f(\mathbf{x}) = \frac{1}{\pi\,|\Sigma|} \exp\{-(\overline{\mathbf{x} - \boldsymbol{\mu}})' \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu})\}. \tag{5.3}$$

Now, consider the signal process $\mathbf{x}(t)$ in (2.1) but assume that the joint density of $\mathbf{x}_1 = \mathbf{x}(t_1), ..., \mathbf{x}_N = \mathbf{x}(t_N)$ is

$$f(\mathbf{x}_1, ..., \mathbf{x}_N) = |\Sigma|^{-N} h(N \operatorname{tr} \Sigma^{-1}\hat{\Sigma}) \tag{5.4}$$

where $N\hat{\Sigma} = \sum_{j=1}^{N} \mathbf{x}_j \bar{\mathbf{x}}_j'$. Let $\lambda_1 \geqslant \cdots \geqslant \lambda_p$ be the eigenvalues of Σ and let $\delta_1 \geqslant \cdots \geqslant \delta_p$ denote the eigenvalues of $\hat{\Sigma}$. Also, let M_k denote the model in which

$$\lambda_1 \geqslant \cdots \geqslant \lambda_k > \lambda_{k+1} = \cdots = \lambda_p = \sigma^2 \tag{5.5}$$

where σ^2 is unknown. Let $f(\mathbf{x}_1, ..., \mathbf{x}_N \mid \theta)$, $\theta \in \Theta_k$, denote the likelihood function under kth model M_k. Also, let

$$L(\theta) = f(\mathbf{x}_1, ..., \mathbf{x}_N \mid \theta), \qquad \theta \in \Theta_k \tag{5.6}$$

for $k = 0, 1, ..., p-1$. We know that for given $\lambda_1, ..., \lambda_p$ the minimum of $\operatorname{tr} \Sigma\hat{\Sigma}^{-1}$ is $\sum_{j=1}^{p} \lambda_i^{-1}\delta_i$ (see [25]). So,

$$\max_{\theta \in \Theta_k} \log L(\theta) = \max_{\theta \in \Theta_k} \left\{ -N \sum_{j=1}^{p} \log \lambda_j + \log h\left(N \sum_{j=1}^{p} \lambda_j^{-1} \delta_j \right) \right\} \tag{5.7}$$

where the maximum is taken subject to (5.5). Suppose $h(t)$ has a continuous derivative $h'(t)$ on $[0, \infty)$ and the equation

$$Nph(y) = y\,|h'(y)| \tag{5.8}$$

has a unique solution $y = Np/\gamma_h$. Then, the above maximum is reached at

$$\frac{\lambda_1}{\delta_1} = \cdots = \frac{\lambda_k}{\delta_k} = \sigma^2 \Big/ \left\{ \frac{1}{p-k} (\delta_{k+1} + \cdots + \delta_p) \right\} = \gamma_h \tag{5.9}$$

and

$$\max_{\theta \in \Theta_k} \log L(\theta) = -Np \log \gamma_h + \log h(Np/\gamma_h) - N \sum_{i=1}^{p} \log \delta_i + G_1(k) \tag{5.10}$$

where

$$G_1(k) = N \log \left[\prod_{i=k+1}^{p} \delta_i \Big/ \left(\frac{1}{p-k} \sum_{i=k+1}^{p} \delta_i \right)^{p-k} \right]. \tag{5.11}$$

So, we can use the procedure discussed in Section 3 to determine the number of signals, even when the observation vectors are jointly distributed as elliptically symmetric. But, we do not know whether this procedure is consistent.

In general, we may use the procedure discussed in Section 3 when σ^2 is unknown even if the observations $\mathbf{x}(t_1),\dots, \mathbf{x}(t_N)$ are not independent and Gaussian provided the conditions on the observations stated in Theorem 3.1 are satisfied. Of course, when the observations are not independent and Gaussian, the statistic L_k in the above procedure is not the likelihood ratio test statistic but the procedure is strongly consistent. As an alternative procedure, we estimate q with $\hat{q}$ where

$$\hat{q} = \max\{k \leqslant p-1 : G_1(k) - G_1(k-1) > C_N\}. \tag{5.12}$$

Under the conditions of Therem 3.1. we observe that, for $k < q$,

$$\lim_{N \to \infty} \frac{1}{N} (G_1(k) - G_1(k-1)) = W(k, k-1) > 0 \qquad \text{a.s.} \tag{5.13}$$

and for $k > q$ $(k \leqslant p-1)$,

$$\frac{1}{N} (G_1(k) - G_1(k-1)) = O\left(\frac{\log \log N}{N} \right) \qquad \text{a.s.} \tag{5.14}$$

where

$$\begin{aligned} &W(k, k-1)\\ &\quad=(p-k+1)\left[\log\left(\frac{1}{(p-k+1)}(\lambda_k+\cdots+\lambda_p)\right.\right.\\ &\quad\quad\left.-\frac{1}{(p-k+1)}\log\lambda_k-\frac{(p-k)}{(p-k+1)}\log\left(\frac{1}{(p-k)}(\lambda_{k+1}+\cdots+\lambda_p)\right)\right]. \end{aligned} \tag{5.15}$$

We see that $G_1(k)$ is non-decreasing function of k for $k \in \{0, 1,..., p-1\}$. If we draw the points $(0, G_1(0))$, $(1, G_1(1)),..., (p-1, G_1(p-1))$ in the Descartes coordinate plane, and construct a polygonal line with these points as its p vertexes, then $G_1(k)-G_1(k-1)$ is just the slope of the kth segment. Suppose that q is the true number of signals. For convenience we temporarily assume $q>0$. As shown in (5.13) and (5.14), we can assert with probability one that, for large N,

$$G_1(k)-G_1(k-1) \geqslant C_1 N \qquad \text{for} \quad k \leqslant q \tag{5.16}$$

and

$$G_1(k)-G_1(k-1)=O(\log\log N) \qquad \text{for} \quad q<k\leqslant p-1, \tag{5.17}$$

where $C_1>0$ is a constant. Thus we see that, the slope $G_1(k)-G_1(k-1)$ has a significant change for $k\leqslant q$ and $q<k\leqslant p-1$, and the true value q is just the largest k for which $G_1(k)-G_1(k-1)>C_N$, where C_N satisfies the following conditions:

$$\lim_{N\to\infty}(C_N/N)=0 \qquad \lim_{N\to\infty}(C_N/\log\log N)=\infty. \tag{5.18}$$

If we put $G_1(-1)=-\infty$, then the same is true for $q=0$. Motivated by (5.16) and (5.17), we estimate the number of signals q with $\hat{q}$ where

$$\hat{q}=\max\{k\leqslant p-1: G_1(k)-G_1(k-1)>C_N\}. \tag{5.19}$$

Under the conditions of Theorem 3.1, we can show that $\hat{q}$ is a consistent estimate of q by following the same lines as in Section 3.

In general, we do not know whether the conditions of Theorem 3.1 are satisfied. In these cases, we make the following assumptions:

$$\begin{aligned} &\text{(i)} \quad \lim_{N\to\infty}\frac{1}{N}(G_1(k)-G_1(k-1)) \overset{\text{a.s.}}{=} 0 && \text{for} \quad k>q\\ &\text{(ii)} \quad \lim_{N\to\infty}\frac{1}{N}(G_1(k)-G_1(k-1)) \overset{\text{a.s.}}{=} W(k,k-1)>0 && \text{for} \quad k\leqslant q \end{aligned} \tag{5.20}$$

where we denote $\lambda_0 = \infty$ for convenience. In this case, we need to assume that the smallest non-zero eigenvalue of $A\psi\bar{A}'$ is distinguishable from σ^2, namely, the ratio of signal intensity to that of noise can be detected by the sensor. We assume that $(\lambda_q - \sigma^2)/\sigma^2 \geqslant \varepsilon > 0$ and ε is known for the given receiver. In this case, we estimate q with $\hat{q}$ where $\hat{q}$ is chosen such that

$$\hat{q} = \max\{k \leqslant p-1 : G_1(k) - G_1(k-1) > \frac{\mu}{2} N\}, \tag{5.21}$$

where we denote $G_1(-1) = -\infty$ for convenience. Also,

$$\begin{aligned}\mu = \min_{0 \leqslant k \leqslant p-1} (p-k+1)\Big\{&\log\left(\frac{1}{p-k+1} + \frac{p-k}{p-k+1}\delta\right) \\ &- \frac{p-k}{p-k+1}\log\delta\Big\} > 0,\end{aligned} \tag{5.22}$$

and

$$\delta = 1 - \frac{\varepsilon}{p(1+\varepsilon)}. \tag{5.23}$$

We now establish the strong consistency of $\hat{q}$. To prove this, we write

$$\alpha_k = \frac{1}{p-k+1}, \qquad \beta_k = \frac{p-k}{p-k+1},$$

$$A_k = \frac{1}{p-k} \sum_{i=k+1}^{p} \lambda_i/\lambda_k, \qquad \lambda_0 = \infty.$$

Suppose that q is the true number of signals and $k \leqslant q$. Then

$$\lim_{N\to\infty} \frac{1}{N}(G_1(k) - G_1(k-1)) = W(k, k-1) \qquad \text{a.s.} \tag{5.24}$$

$$W(k, k-1) = (p-k+1)\{\log(\alpha_k + \beta_k A_k) - \beta_k \log A_k\} > 0.$$

Consider $f_k(x) = \log(\alpha_k + \beta_k x) - \beta_k \log x$ for $x \in (0, 1]$. We have

$$f_k'(x) = -\alpha_k\beta_k(1-x)/x(\alpha_k + \beta_k x) < 0, \qquad 0 < x \leqslant 1,$$

so that $f_k(x)$ is a decreasing function on $(0, 1]$. But if $\lambda_q \geqslant (1+\varepsilon)\,\sigma^2$, then for $0 \leqslant k \leqslant q-1$,

$$A_k \leqslant \frac{p-q}{p-k}\frac{1}{1+\varepsilon} + \frac{q-k}{p-k} \leqslant 1 - \frac{1}{p-k}\frac{\varepsilon}{1+\varepsilon} \leqslant 1 - \frac{\varepsilon}{p(1+\varepsilon)} = \delta$$

and

$$A_q = \sigma^2/\lambda_q \leqslant \frac{1}{1+\varepsilon} < \delta.$$

Thus for $0 \leqslant k \leqslant q$,

$$W(k, k-1) \geqslant \mu. \tag{5.25}$$

From (5.24) and (5.25), it follows that, with probability one for large N,

$$G_1(k) - G_1(k-1) > \frac{\mu}{2} N, \qquad k \leqslant q. \tag{5.26}$$

On the other hand, if $q < k \leqslant p-1$,

$$\lim_{N \to \infty} \frac{1}{N} (G_1(k)) - G_1(k-1)) = 0 \qquad \text{a.s.} \tag{5.27}$$

So with probability one, for large N,

$$G_1(k) - G_1(k-1) < \frac{\mu}{2} N \qquad \text{for} \quad q < k \leqslant p-1. \tag{5.28}$$

Thus from (5.26) and (5.28) it follows, with probability one, for large N,

$$\hat{q} = q. \tag{5.29}$$

and the assertion is proved.

6. Inference on the Multiplicities of the Eigenvalues of Σ When σ^2 Is Unknown

In Section 3, we discussed the problem of determination of the number of signals when σ^2 is unknown. As pointed out earlier, this problem is equivalent to drawing inference on the multiplicity of the smallest eigenvalue. In this section, we discuss the problem of not only finding the number of signals but also determination of the multiplicities of the eigenvalues of Σ which are not equal to the smallest eigenvalue. For the integer interval $[0, p]$ there exist 2^{p-1} different integer partitions such as $0 = k_0 < \cdots < k_l = p$, $l = 1, 2, \ldots, p$. Let $(k_1, \ldots, k_l)$ denote such a partition and let $\mathscr{K}$ be the set of all of these partitions. Let $M_{k_1 \cdots k_l}$ denote the model for which the eigenvalues $\lambda_1, \ldots, \lambda_p$ have the multiplicities as denoted

$$\lambda_{k_{i-1}+1} = \cdots = \lambda_{k_i}, \qquad i = 1, 2, \ldots, l \tag{6.1}$$

and $\lambda_{k_1} > \lambda_{k_2} > \cdots > \lambda_{k_l}$. We are interested in selecting one of the 2^{p-1} possible models $M_{k_1 \cdots k_l}$. Let $\Theta_{k_1 \cdots k_l}$ denote the parametric space for which (6.1) is true. When the observations are distributed independently as multivariate normal, the logarithm of the likelihood function under (6.1) is

$$\log L(\theta) = -\frac{N}{2} \log |\Sigma| - \frac{N}{2} \operatorname{tr} \Sigma^{-1} \hat{\Sigma}, \qquad \theta \in \Theta_{k_1 \cdots k_l} \tag{6.2}$$

and σ^2 is unknown. It is known (see [2]) that

$$\sup_{\theta \in \Theta_{k_1 \cdots k_l}} \log L(\theta) = -\frac{N}{2} \sum_{i=1}^{l} \Delta k_i \log \left(\frac{1}{\Delta k_i} \sum_{j=k_{i-1}+1}^{k_i} \delta_j \right) - \frac{Np}{2}. \tag{6.3}$$

where $\delta_1 \geqslant \cdots \geqslant \delta_p$ are the eigenvalues of $\hat{\Sigma}$, and $\Delta k_i = k_i - k_{i-1}$. Now, let

$$L(k_1, ..., k_l) = -\frac{1}{2} N \sum_{i=1}^{l} \Delta k_i \log \left(\frac{1}{\Delta k_i} \sum_{j=k_{i-1}+1}^{k_i} \delta_j \right), \tag{6.4}$$

and

$$G(k_1, ..., k_l) = L(k_1, ..., k_l) - lC_N \tag{6.5}$$

where C_N satisfies the following conditions:

$$\lim_{N \to \infty} (C_N / N) = 0 \qquad \text{and} \qquad \lim_{N \to \infty} (C_N / \log \log N) = \infty. \tag{6.6}$$

Then, we can find the partition $(\hat{q}_1, ..., \hat{q}_{\hat{r}}) \in \mathscr{K}$ which satisfies

$$G(\hat{q}_1, ..., \hat{q}_{\hat{r}}) = \max_{(k_1, ..., k_l) \in \mathscr{K}} G(k_1, ..., k_l) \tag{6.7}$$

and use $(\hat{r}, \hat{q}_1, ..., \hat{q}_{\hat{r}})$ as an estimate of the true partition $(r, q_1, ..., q_r)$ corresponding to the true model. We now prove the following theorem:

THEOREM 6.1. *If $N\hat{\Sigma} \sim W_p(N, \Sigma)$, then, $(\hat{r}, \hat{q}_1, ..., \hat{q}_{\hat{r}})$ is a strongly consistent estimate of $(r, q_1, ..., q_r)$ when $M_{q_1 \cdots q_r}$ is the true model.*

Remark 1. For the strong consistency of $(\hat{r}, \hat{q}_1, ..., \hat{q}_{\hat{r}})$ we need not assume that $N\hat{\Sigma} \sim W_p(N, \Sigma)$. We can use $G(k_1, ..., k_r)$ in (6.5) even in the case when $L(k_1, ..., k_r)$ is not necessarily a likelihood ratio test statistic. We assume that $N\hat{\Sigma} = \sum_{i=1}^{N} \mathbf{x}_i \bar{\mathbf{x}}_i'$, $E\mathbf{x}_i = \mathbf{0}$, $E(\mathbf{x}_i \bar{\mathbf{x}}_i') = \Sigma > 0$, $E(\bar{\mathbf{x}}_i' \mathbf{x}_i)^2 < \infty$, and $\mathbf{x}_1, ..., \mathbf{x}_n$ are i.i.d. Then $(\hat{r}, \hat{q}_1, ..., \hat{q}_{\hat{r}})$ is still a strongly consistent estimate.

Proof. Suppose that $M_{q_1, ..., q_r}$ is the true model and $\lambda_1 \geqslant \cdots \geqslant \lambda_p$ are the eigenvalues of Σ. By the law of the iterated logarithm, we have

$$\hat{\Sigma} - \Sigma = O(\sqrt{\log \log N / N}) \qquad \text{a.s.} \quad \text{as} \quad N \to \infty.$$

By Lemma 3.2, we have

$$|\delta_i - \lambda_i| = O\left(\left(\frac{1}{N}\log\log N\right)^{1/2}\right) \qquad \text{a.s.} \tag{6.8}$$

as $N \to \infty$ for $i = 1,..., p$.

Now, we give a definition of the refinement of a partition. A partition $(k_1,..., k_l) \in \mathscr{K}$ is called a refinement of the partition $(j_1,..., j_m) \in \mathscr{K}$, if $\{j_1,..., j_m\}$ is a proper subset of $\{k_1,..., k_l\}$. This fact is written as $(k_1,..., k_l) \prec (j_1,..., j_m)$. Suppose that $(k_1,..., k_{i-1}, k_i', k_i,..., k_l) \prec (k_1,..., k_l)$. Then, by Jensen's inequality, we have

$$\begin{aligned} &L(k_1,..., k_{i-1}, k_i', k_i,..., k_l) - L(k_1,..., k_l) \\ &\quad = \tfrac{1}{2}N\Delta k_i[-\log(p_1 A_1 + p_2 A_2) + p_1 \log A_1 + p_2 \log A_2] \geqslant 0 \end{aligned} \tag{6.9}$$

where

$$A_1 = \frac{1}{k_i' - k_{i-1}} \sum_{j=k_{i-1}+1}^{k_i'} \delta_j, \qquad A_2 = \frac{1}{k_i - k_i'} \sum_{j=k_i'+1}^{k_i} \delta_j,$$

$$p_1 = \frac{k_i' - k_{i-1}}{\Delta k_i}, \qquad p_2 = \frac{k_i - k_i'}{\Delta k_i}.$$

Thus, if $(k_1,..., k_l) \prec (j_1,..., j_m)$, then

$$L(j_1,..., j_m) \leqslant L(k_1,..., k_l). \tag{6.10}$$

Now, we assume that $(k_1,..., k_l) \prec (q_1,..., q_r)$. Then $l > r$. Next, write $\kappa_i = \{q_{i-1}+1, q_{i-1}+2,..., q_i\}$, $\Delta q_i = a_i$, and $\mu_i = (\delta_i - \lambda_i)/\lambda_i$. Then we have

$$\begin{aligned} 0 &\leqslant L(k_1,..., k_l) - L(q_1,..., q_r) \leqslant L(1, 2,..., p) - L(q_1,..., q_r) \\ &= \frac{1}{2} N \sum_{i=1}^{r} \left[a_i \log\left(1 + \frac{1}{a_i}\sum_{j\in\kappa_i} \mu_j\right) - \sum_{j\in\kappa_i} \log(1+\mu_j)\right]. \end{aligned} \tag{6.11}$$

By (6.8) we have $\lim_{N\to\infty} \mu_i = 0$ a.s. Using Taylor's expansion, we have

$$\begin{aligned} &0 \leqslant L(1,..., p) - L(q_1,..., q_r) \\ &= \frac{1}{4} N \sum_{i=1}^{r} \left[\sum_{j\in\kappa_i} \mu_j^2 (1+o(1)) - \frac{1}{a_i}\left(\sum_{j\in\kappa_i}\mu_j\right)^2 (1+o(1))\right] \qquad \text{a.s.} \end{aligned} \tag{6.12}$$

By (6.8) we have $\mu_j = O((\log\log N/N)^{1/2})$, a.s. and so,

$$0 \leqslant L(1,..., p) - L(q_1,..., q_r) = O(\log\log N) \qquad \text{a.s.} \tag{6.13}$$

Using (6.5), (6.6), (6.11), and (6.13), we have with probability one for large N

$$\begin{aligned} &G(q_1,..., q_r) - G(k_1,..., k_l) \\ &\qquad = L(q_1,..., q_r) - L(k_1,..., k_l) + (l-r)\, C_N > 0. \end{aligned} \tag{6.14}$$

Now we assume that $(k_1,..., k_l)$ is a partition of $[0, p]$ such that there exists at least one q_t satisfying $k_{i-1} < q_t < k_i$ for some i. Let $N_i = \{k_{i-1}+1,..., k_i\}$. Using (6.8) and (6.13), we obtain

$$\begin{aligned} &\lim_{N\to\infty} \frac{1}{N} [L(q_1,..., q_r) - L(k_1,..., k_l)] \\ &\quad = \lim_{N\to\infty} \frac{1}{N} [-L(1, 2,..., p) + L(q_1,..., q_r)] \\ &\quad\quad + \lim_{N\to\infty} \frac{1}{N} [L(1,..., p) - L(k_1,..., k_l)] \qquad (6.15) \\ &\quad = \lim_{N\to\infty} \frac{1}{4} \sum_{i=1}^{r} \Delta k_i \left[-\frac{1}{\Delta k_i} \sum_{j\in N_i} \log \delta_j + \log\left(\frac{1}{\Delta k_i} \sum_{j\in N_i} \delta_j \right) \right] \\ &\quad = \frac{1}{4} \sum_{i=1}^{r} \Delta k_i \left[-\frac{1}{\Delta k_i} \sum_{j\in N_i} \log \lambda_j + \log\left(\frac{1}{\Delta k_i} \sum_{j\in N_i} \lambda_j \right) \right] \qquad \text{a.s.} \end{aligned}$$

By Jensen's inequality, we have

$$-\frac{1}{\Delta k_i} \sum_{j\in N_i} \log \lambda_j + \log\left(\frac{1}{\Delta k_i} \sum_{j\in N_i} \lambda_j \right) \geqslant 0, \qquad i = 1,..., r, \tag{6.16}$$

and there exists at least one i such that $\lambda_j, j \in N_i$ are not all identical. So the inequalities in (6.16) are strict for some i. Thus, with probability one for large N, we have

$$L(q_1,..., q_r) - L(k_1,..., k_l) \geqslant cN \tag{6.17}$$

where $c > 0$ is a constant. From (6.17) and $\lim_{N\to\infty} (C_N/N) = 0$, it follows that, with probability one for large N,

$$G(q_1,..., q_r) - G(k_1,..., k_l) > 0. \tag{6.18}$$

By (6.14) and (6.18), with probability one for large N,

$$\hat{r} = r \quad (\hat{q}_1,..., \hat{q}_{\hat{r}}) = (q_1,..., q_r). \tag{6.19}$$

This completes the proof.

7. Inference on the Multiplicities of the Eigenvalues of Σ When σ^2 Is Known

In this section, we discuss the problem of determination of the multiplicities of the eigenvalues when σ^2 is known. Without loss of generality, we assume that $\sigma^2=1$. Let $M^*_{k_1 \dots k_l}$ denote the model for which the eigenvalues have multiplicities as given below

$$H^*(\mu_1,\dots,\mu_l):\begin{cases}\lambda_{k_{i-1}+1}=\cdots=\lambda_{k_i}=\tilde{\lambda}_i,\ i=1,2,\dots,l-1,\\ \lambda_{k_{l-1}+1}=\cdots=\lambda_p=1\end{cases} \tag{7.1}$$

where $k_0=0$, $k_h=k_{h-1}+\mu_h (h=1,2,\dots,l)$. We are interested in selecting one of the 2^{p-1} possible models $M^*_{k_1 \cdots k_l}$. Let $\Theta^*_{k_1 \cdots k_l}$ denote the parametric space for which (7.1) is true. Also, let $L^*(\theta)$ denote the likelihood function when (7.1) is true, that is,

$$\log L^*(\boldsymbol{\theta})=-\frac{N}{2}\log|\Sigma|-\frac{N}{2}\operatorname{tr}\Sigma^{-1}\hat{\Sigma},\qquad \boldsymbol{\theta}\in\Theta^*_{k_1\cdots k_l}. \tag{7.2}$$

By von Neumann's inequality, we have

$$\operatorname{tr}\Sigma^{-1}\hat{\Sigma}\leqslant\sum_{i=1}^{p}(\delta_i/\lambda_i)=\sum_{h=1}^{l-1}\frac{1}{\tilde{\lambda}_h}\sum_h\delta_i+\sum_l\delta_i \tag{7.3}$$

where the summation $\sum_h$ is carried over $i=k_{h-1}+1,\dots,k_h$. So,

$$\begin{aligned}L^*(\mu_1,\dots,\mu_l)&=\sup_{\boldsymbol{\theta}\in\Theta^*_{\mu_1\cdots\mu_l}}\log L^*(\boldsymbol{\theta})\\&=-\frac{N}{2}\left\{\inf_{\tilde{\lambda}_1>\cdots>\tilde{\lambda}_{l-1}>1}\sum_{h=1}^{l-1}\left(\mu_h\log\tilde{\lambda}_h+\frac{1}{\tilde{\lambda}_h}\sum_h\delta_i\right)\right\}.\end{aligned} \tag{7.4}$$

Let $\tau=\max\{h\leqslant l:\sum_h\delta_i\geqslant\mu_h\}$. If $h\leqslant\tau$ then we can easily know by differentiation

$$\inf_{\tilde{\lambda}_h>1}\left(\frac{1}{\tilde{\lambda}_h}\sum_h\delta_i+\mu_h\log\tilde{\lambda}_h\right)=\mu_h\log\left(\frac{1}{\mu_h}\sum_h\delta_i\right)+\mu_h. \tag{7.5}$$

If $h>\tau$, $\sum_h\delta_i<\mu_h$, the only point $\tilde{\lambda}_h=(1/\mu_h)\sum_h\delta_i$ which minimizes $(1/\tilde{\lambda}_h)\sum_h\delta_i+\mu_h\log\tilde{\lambda}_h$ is located outside the region $\tilde{\lambda}_h>1$, and we get that

$$\inf_{\tilde{\lambda}_h>1}\left(\frac{1}{\tilde{\lambda}_h}\sum_h\delta_i+\mu_h\log\tilde{\lambda}_h\right)=\sum_h\delta_i. \tag{7.6}$$

From (7.4)–(7.6) and the fact that

$$\frac{1}{\mu_1}\sum_1 \delta_i \geqslant \frac{1}{\mu_2}\sum_2 \delta_i \geqslant \cdots \geqslant \frac{1}{\mu_\tau}\sum_\tau \delta_i \geqslant 1,$$

we obtain

$$L^*(\mu_1,..., \mu_l) = -\frac{N}{2}\left\{\sum_{h=1}^{\tau} \mu_h \log \frac{1}{\mu_h}\sum_h \delta_i + k_\tau + \sum_{i=k_\tau+1}^{p} \delta_i\right\}. \tag{7.7}$$

It is well known that the maximum of the log-likelihood under no restrictions on the parametric space is given by

$$-\frac{N}{2}\left\{p + \sum_{i=1}^{p} \log \delta_i\right\} \tag{7.8}$$

From (7.7) and (7.8) we get the logarithm of the likelihood ratio test statistics under $H^*(\mu_1,..., \mu_l)$ which is given by

$$\begin{aligned} G^*(\mu_1,..., \mu_l) = &-\frac{N}{2}\sum_{h=1}^{\tau} \mu_h \left(\log \frac{1}{\mu_h}\sum_h \delta_i - \frac{1}{\mu_h}\sum_h \log \delta_i\right) \\ &-\frac{N}{2}\sum_{i=k_\tau+1}^{p} (1 + \delta_i - \log \delta_i). \end{aligned} \tag{7.9}$$

In the following we consider $\mathbf{x}_1,..., \mathbf{x}_n$ as identically distributed random p-vectors with mean zero and covariance matrix $\Sigma = Q \operatorname{diag}[\tilde{\lambda}_1 I_{\mu_1},..., \tilde{\lambda}_{l-1} I_{\mu_{l-1}}, I_{\mu_l}]\, Q'$ with Q an orthogonal matrix and we do not assume $\mathbf{x}_i$'s are independent and multivariate normal. Of course, in general we cannot get the logarithm of likelihood ratio test as given in (7.9). But we can regard $G^*(\mu_1,..., \mu_l)$ as a statistic for testing the hypothesis $H^*(\mu_1,..., \mu_l)$. Now, our purpose is to find an information criterion for the detection of the multiplicities $\mu_1,..., \mu_l$. Define

$$I^*(\mu_1,..., \mu_l; C_N) = -2G^*(\mu_1,..., \mu_l) + lC_N$$

where C_N satisfies

(1) $C_N/N \to 0$.

(2) $C_N/\log\log N \to \infty$

and determine $\hat{\mu}_1,..., \hat{\mu}_{\hat{l}}$ satisfying

$$I^*(\hat{\mu}_1,..., \hat{\mu}_{\hat{l}}; C_N) = \min\left\{I^*(\mu_1,..., \mu_l; C_N)\colon \sum_{h=1}^{l} \mu_h = p,\ l = 1, 2,..., p\right\}.$$

As before, we can similarily prove the following theorem.

THEOREM. *Suppose that* $\mathbf{x}_1, \mathbf{x}_2, \ldots,$ *is a complex or real, stationary, and* ϕ*-mixing process with* ϕ *being decreasing and such that* $\sum_{h=1}^{\infty} \phi^{1/2}(h) < \infty$. *Also* $E(\bar{\mathbf{x}}_i' \mathbf{x}_i)^2 < \infty$. Then

$$\{\hat{l}, \hat{\mu}_1, \ldots, \hat{\mu}_l\} \to \{l, \mu_1, \ldots, \mu_l\} \qquad \text{a.s.} \tag{7.10}$$

Under the conditions of the theorem, with probability one, when N is large enough, $\tau \geqslant l - 1$. Then the proof of the theorem is essentially the same as that for λ_l unknown and we omit it.

REFERENCES

[1] AKAIKE, H. (1972). Information theory and an extension of the maximum likelihood principle. In *Proceedings of the Second International Symposium on Information Theory, Supp. to Problems of Control and Information Theory*, pp. 267–281.

[2] ANDERSON, T. W. (1963). Asymptotic theory for principal component analysis. *Ann. of Math. Statist.* **34** 122–148.

[3] BAI, Z. D. (1984). *A Note on Asymptotic Joint Distribution of the Eigenvalues of a Noncentral Multivariate F Matrix*. Tech. Rept. No. 84-49. Center for Multivariate Analysis, University of Pittsburgh.

[4] BAI, Z. D., KRISHNAIAH, P. R., AND ZHAO, L. C. (1985). On Rates of Convergence of Efficient Detection Criteria in Signal Processing with White Noise. Tech. Rept. No. 85-45. Center for Multivariate Analysis, University of Pittsburgh.

[5] BARTLETT, M. S. (1954). A note on the multiplying factors for various χ^2 approximations. *J. R. Statist. Soc. Ser. B* **16** 296–298.

[6] GOODMAN, N. R. (1963). Statistical analysis based on a certain multivariate complex Gaussian distribution (an introduction). *Ann. of Math. Statist.* **34** 152–176.

[7] HALL, P., AND HEYDE, C. C. (1980). *Martingale Limit Theory and Its Application.* Academic Press, New York.

[8] HANNAN, E. J., AND QUINN, B. G. (1979). The determination of the order of an autoregression. *J. R. Statist. Soc. Ser. B* **41** 190–195.

[9] KELKER, D. (1970). Distribution theory of spherical distribution and a location scale parameter generalization. *Sankhya Ser. A* **43** 419–430.

[10] KRISHNAIAH, P. R. (1976). Some recent developments on complex multivariate distributions. *J. Multivariate Anal.* **6** 1–30.

[11] KRISHNAIAH, P. R. (1978). Some developments on real multivariate distributions. In *Developments in Statistics*, Vol. 1 (P. R. Krishnaiah, Ed.), pp. 135–169, Academic Press, New York.

[12] KRISHNAIAH, P. R., AND LIN, J. (in press). Complex elliptical distributons. *Commun. in Statist.*

[13] KRISHNAIAH, P. R., AND SCHUURMANN, F. J. (in preparation). *Computations of Complex Multivariate Distributions.* To be published by North-Holland, Amsterdam/New York.

[14] KRISHNAIAH, P. R., AND WAIKAR, V. B. (1971). Simultaneous tests for equality of latent roots against certain alternatives, I. *Ann. Inst. Statist. Math.* **24** 81–85.

[15] KRISHNAIAH, P. R., AND WAIKAR, V. B. (1972). Simultaneous tests for equality of latent roots against certain alternatives, II. *Ann. Inst. Statist. Math.* **24** 81–85.

[16] LIGGETT, W. S. (1973). Passive sonar: Fitting models to multiple time series, In *Signal Processing* (J. W. Griffith *et al.*, Eds.), pp. 327–345. Academic Press, New York.

[17] RAO, C. R. (1973). *Linear Statistical Inference and Its Applications.* Wiley, New York.

[18] RAO, C. R. (1983). Likelihood ratio tests for relationships between two covariance matrices. In *Studies in Econometrics, Time Series and Multivariate Statistics* (T. Amemiya, S. Karlin, and L. Goodman, Eds.). Academic Press, New York.

[19] REZNIK, M. K. (1968). The law of the iterated logarithm for some classes of stationary processes. *Theory Probab. Appl.* **8** 606–621.

[20] RISSANEN, J. (1978). Modeling by shortest data description. *Automatica,* **14** 465–471.

[21] SCHMIDT, R. O. (1979). Multiple emitter location and signal parameter estimation. In *Proc. RADC Spectrum Estimation Workshop,* Rome.

[22] SCHWARTZ, G. (1978). Estimating the dimension of a model. *Ann. Statist.* **6** 461–464.

[23] STOUT, W. F. (1974). *Almost Sure Convergence.* Academic Press, New York.

[24] TUFTS, D. W., AND KUMARESAN, R. (1980). Data-adaptive principal component signal processing. *Proceedings, 19th IEEE Conf. Decision Control,* Albuquerque, N.M., pp. 949–954.

[25] VON NEUMANN, J. (1937). Some matrix-inequalities and metrization of metric space. *Tomsk Univ. Rev.* **1** 286–300.

[26] WAX, M., AND KAILATH, T. (1984). Determination of the number of signals by information theoretic criteria. *IEEE Trans. Acoustics Speech Signal Process.* **ASSP-33**.

[27] WAX, H., SHAN, T. J., AND KAILATH, T. (1984). Spatio-temporal spectral analysis by eigenstructure methods. *IEEE Trans. Acoustics Speech Signal Process.* **ASSP-32** 817–827.

[28] WHALEN, A. D. (1971). *Detection of Signals in Noise.* Academic Press, New York.

[29] WOODING, R. A. (1956). The multivariate distribution of complex normal variables. *Biometrika* **43** 212–215.

[30] ZHAO, L. C., KRISHNAIAH, P. R., AND BAI, Z. D., (in press). Remarks on certain criteria for detection of number of signals. *IEEE Trans. Acoustics Speech Signal Process.*

[31] ZHAO, L. C., KRISHNAIAH, P. R. AND BAI, Z. D. (1986). On detection of the number of signals when the noise covariance matrix is arbitrary. *J. Multivariate Anal.* **20** 26–49.