ISSN:1991-8178

Australian Journal of Basic and Applied Sciences

Journal home page: www.ajbasweb.com

Consistency and Efficiency of the Maximum Likelihood Method of Gaussian Observations Not Similarly Distributed

[1]Hussam Kassem, [1]Hossein Nassar and [2]Hamad Saber

[1] *Lebanese University, Faculty of Public Health, P.O. Box 8443. . Karak, Bekaa, Lebanon.*
[2]*Lebanese International University, P. O. Box 5, Jeb Janeen, Al-Khyara, Bekaa, Lebanon.*

ARTICLE INFO

Article history:
Received 12 March 2015
Accepted 28 April 2015
Available online 5 May 2015

Keywords:
Maximum Likelihood Estimator, Consistency, Efficiency, Gaussian Observations.

ABSTRACT

Most results concerning Consistency and Efficiency of the Maximum Likelihood (ML) method have been established in the case of independent observations distributed according to the **same** probability density function [Cramer, 1949]. The purpose of this paper is to give sufficient conditions to show the ML is consistent and efficient in the case of independent Gaussian observations that are **not** distributed according to the same law.

© 2015 AENSI Publisher All rights reserved.

To Cite This Article: Hussam Kassem, Hossein Nassar and Hamad Saber., Consistency and Efficiency of the Maximum Likelihood Method of Gaussian Observations Not Similarly Distributed. ***Aust. J. Basic & Appl. Sci.,*** *9(11): 14-23, 2015*

1. INTRODUCTION

In active communications systems, the primary goal is to detect the presence and estimate the parameters of signal in the presence of noise and jammers. For radio mobile cellular system, the parameters of interest include the signal amplitude, direction of arrival, delays and the covariance matrix of noise and jammers.

Standard solutions to this problem involve the use of classical space-time filters. Maximum Likelihood approaches have not been established in the case of independent Gaussian observations not distributed according to the same law. We apply the last method to array processing, and we study the asymptotic statistical properties of the ML Estimator (MLE).

These ML approaches demonstrate that MLE is consistent and efficient taking into account some hypothesizes.

Section 2 reminds the data model, and presents the consistency, the efficiency of the MLE, and calculates the corresponding Fisher Information Matrix (FIM), then section 3 concludes.

2. CONSISTENCY AND EFFICIENCY

2.1 Data Model:

Let us consider the following model that is commonly met in array processing:

$$\mathbf{y}(t) = \mathbf{s}(t, \boldsymbol{\theta}_0) + \mathbf{n}(t) \quad \text{for } t = 1, \cdots, N \tag{1}$$

where:

- $\mathbf{y}(t)$ is the noisy measured signal with M components;
- $\mathbf{s}(t, \boldsymbol{\theta})$ is a deterministic information bearing signal that depends on an unknown vector parameter $\boldsymbol{\theta} = \left[\theta_1 \quad \cdots \quad \theta_{P_\theta}\right]^T$ whose exact value is $\boldsymbol{\theta}_0$;
- $\mathbf{n}(t)$ is a temporally white complex Gaussian, circular, zero-mean noise with covariance matrix $\mathbf{Q}(\boldsymbol{\mu})$, where $\boldsymbol{\mu} = \left[\mu_1 \quad \cdots \quad \mu_{P_\mu}\right]^T$ is an unknown vector parameter.

Thus, observations $\mathbf{y}(t)$ have a common covariance matrix **but a time-varying mean**. Note that model (1) is met in a lot of practical situations: in active RADAR, SONAR and channel estimation with a training sequence in radio communications, we have $\mathbf{s}(t, \boldsymbol{\theta}) = \sum_{k=1}^{J} \beta_k \, \delta(t - \tau_k) \, \mathbf{a}(\alpha_k)$ where $\delta(t)$ is a known signal, and α_k, β_k, τ_k are unknown multipath parameters (angle of arrival, complex amplitude and delay).

In the following we set

- $\boldsymbol{\xi} = \left[\boldsymbol{\theta}^T, \boldsymbol{\mu}^T\right]^T$, $\boldsymbol{\xi}_0$, $\boldsymbol{\theta}_0$ and $\boldsymbol{\mu}_0$ the true values of the parameters ;
- $\mathbf{Y} = \left[\mathbf{y}(1) \quad \cdots \quad \mathbf{y}(N)\right]$;
- $\mathbf{S}(\boldsymbol{\theta}) = \left[\mathbf{s}(1, \boldsymbol{\theta}) \quad \cdots \quad \mathbf{s}(N, \boldsymbol{\theta})\right]$ the signal matrix ;
- $\mathbf{N} = \left[\mathbf{n}(1) \quad \cdots \quad \mathbf{n}(N)\right]$;
- $\mathbf{Q}_0 = \mathbf{Q}(\boldsymbol{\mu}_0)$ and $\mathbf{S}_0 = \mathbf{S}(\boldsymbol{\theta}_0)$.

Corresponding Author: Hussam Kassem, Lebanese University, Faculty of Public Health, P.O. Box 8443. Karak, Bekaa, Lebanon. Tel: +9618930913/14, E-mail: hussamkassem@yahoo.com

Unless necessary, the dependence of **S** and **Q** on $\boldsymbol{\theta}$ and $\boldsymbol{\mu}$ will be omitted. Thus model (1) is equivalent to:

$$\mathbf{Y} = \mathbf{S}_0 + \mathbf{N}. \quad (2)$$

Our purpose is to establish in our context the asymptotic statistical properties (when *N* tends to infinity) of the MLE which is obtained by minimizing the following criterion:

$$L(\boldsymbol{\xi}) = \ln|\mathbf{Q}| + N^{-1} tr\left(\mathbf{Q}^{-1}(\mathbf{Y}-\mathbf{S})(\mathbf{Y}-\mathbf{S})^H\right) \quad (3)$$

where $|\ \ |$ stands for determinant, *tr*( ) denotes the trace and $''^H''$ the conjugate transpose. This problem is solved in this paper when the signal matrix $\mathbf{S}(\boldsymbol{\theta})$ has the following model:

$$\mathbf{S}(\boldsymbol{\theta}) = \mathbf{H}(\boldsymbol{\theta})\,\mathbf{X} \quad (3.a)$$

where:

- $\mathbf{H}(\boldsymbol{\theta})$ is an *unknown Mxq* matrix ;
- **X** is a *known Lxq* matrix.

We also set:

$$\mathbf{H}(\boldsymbol{\theta}_0) = \mathbf{H}_0, \quad (3.b)$$

$$\mathbf{H}(\boldsymbol{\theta}) = \mathbf{H}. \quad (3.c)$$

2.2 Consistency:

Let us consider the following hypothesizes:

- **H1:** $\left\|\mathbf{Q}(\boldsymbol{\mu})^{-1}\right\|_F$ is bounded for all values of $\boldsymbol{\mu}$.
- **H2:** $\|\mathbf{H}\|_F$ is bounded for all values of $\boldsymbol{\theta}$.
- **H3:** $N^{-1}\mathbf{X}\mathbf{X}^H \longrightarrow \mathbf{R}_{xx}$ as $N \longrightarrow \infty$.
- **H4:** $(\mathbf{H}-\mathbf{H}_0)\,\mathbf{R}_{xx}\,(\mathbf{H}-\mathbf{H}_0)^H = \mathbf{0}$ iff $\boldsymbol{\theta} = \boldsymbol{\theta}_0$.

Lemma 1:

Under hypothesizes H1 to H3, the ML criterion (3) converges **with probability one** (w.p.1) uniformly in $\boldsymbol{\xi}$ to the function

$$f(\boldsymbol{\xi}) = \ln|\mathbf{Q}| + tr(\mathbf{Q}^{-1}\mathbf{Q}_0) + tr\left(\mathbf{Q}^{-1}(\mathbf{H}-\mathbf{H}_0)\,\mathbf{R}_{xx}(\mathbf{H}-\mathbf{H}_0)^H\right) \quad (4)$$

Proof: The ML criterion (3) can be rewritten

$$\begin{aligned} L(\boldsymbol{\xi}) &= \ln|\mathbf{Q}| + N^{-1} tr\left(\mathbf{Q}^{-1}(\mathbf{Y}-\mathbf{S})(\mathbf{Y}-\mathbf{S})^H\right) \\ &= \ln|\mathbf{Q}| + N^{-1} tr(\mathbf{Q}^{-1}\mathbf{N}\mathbf{N}^H) \\ &\quad + tr\left(\mathbf{Q}^{-1}(\mathbf{H}-\mathbf{H}_0)(N^{-1}\mathbf{X}\mathbf{X}^H)(\mathbf{H}-\mathbf{H}_0)^H\right) \\ &\quad - 2\Re\left(tr(\mathbf{Q}^{-1}(\mathbf{H}-\mathbf{H}_0)(N^{-1}\mathbf{X}\mathbf{N}^H))\right) \end{aligned}$$

Assume,

$$\begin{aligned} \varepsilon(\boldsymbol{\xi}) &= L(\boldsymbol{\xi}) - f(\boldsymbol{\xi}) \\ &= \varepsilon_1(\boldsymbol{\xi}) + \varepsilon_2(\boldsymbol{\xi}) + \varepsilon_3(\boldsymbol{\xi}) \end{aligned} \quad (5)$$

where:

$$\varepsilon_1(\boldsymbol{\xi}) = tr\left(\mathbf{Q}^{-1}(N^{-1}\mathbf{N}\mathbf{N}^H - \mathbf{Q}_0)\right),$$

$$\varepsilon_2(\boldsymbol{\xi}) = tr\left(\mathbf{Q}^{-1}(\mathbf{H}-\mathbf{H}_0)\left(N^{-1}\mathbf{X}\mathbf{X}^H - \mathbf{R}_{xx}\right)(\mathbf{H}-\mathbf{H}_0)^H\right),$$

$$\varepsilon_3(\boldsymbol{\xi}) = -2\Re\left(tr(\mathbf{Q}^{-1}(\mathbf{H}-\mathbf{H}_0)(N^{-1}\mathbf{X}\mathbf{N}^H))\right).$$

Let us show that $\varepsilon_1(\boldsymbol{\xi})$, $\varepsilon_2(\boldsymbol{\xi})$ and $\varepsilon_3(\boldsymbol{\xi})$ converge w.p.1 to 0 uniformly in $\boldsymbol{\xi}$. Indeed, we have from the Schwartz inequality:

$$|\varepsilon_1(\boldsymbol{\xi})| \le \left\|\mathbf{Q}^{-1}\right\|_F \left\|N^{-1}\mathbf{N}\mathbf{N}^H - \mathbf{Q}_0\right\|_F.$$

From the weak law of large numbers, $\left\|N^{-1}\mathbf{B}\mathbf{B}^H - \mathbf{Q}_0\right\|_F$ converges w.p.1 to 0. Since $\left\|\mathbf{Q}^{-1}\right\|_F$ is bounded by H1, it follows that $\varepsilon_1(\boldsymbol{\xi})$ converges w.p.1 to 0 uniformly in $\boldsymbol{\xi}$.

Similarly,

$$|\varepsilon_2(\boldsymbol{\xi})| \le \underbrace{\left\|(\mathbf{H}-\mathbf{H}_0)^H\mathbf{Q}^{-1}(\mathbf{H}-\mathbf{H}_0)\right\|_F}_{\text{bounded from H1 and H2}} \left\|(N^{-1}\mathbf{X}\mathbf{X}^H - \mathbf{R}_{xx})\right\|_F$$

The right hand side term of this inequality converges uniformly to 0 from H1, H2 and H3. Finally, it is easy to show that $\varepsilon_3(\boldsymbol{\xi})$ is zero-mean Gaussian with variance

$$\sigma^2_{\varepsilon_3} = 2N^{-1} tr\begin{bmatrix} \mathbf{Q}^{-1}\mathbf{Q}_0\mathbf{Q}^{-1}(\mathbf{H}-\mathbf{H}_0) \\ *\,(N^{-1}\mathbf{X}\mathbf{X}^H)\,(\mathbf{H}-\mathbf{H}_0)^H \end{bmatrix} \quad (6)$$

From H1 to H3, the term $tr\left[\mathbf{Q}^{-1}\mathbf{Q}_0\mathbf{Q}^{-1}(\mathbf{H}-\mathbf{H}_0)(N^{-1}\mathbf{X}\mathbf{X}^H)(\mathbf{H}-\mathbf{H}_0)^H\right]$ in (6) is bounded, and thus $\sigma^2_{\varepsilon_3}$ converges to 0 uniformly in $\boldsymbol{\xi}$. Therefore $\varepsilon_3(\boldsymbol{\xi})$ converges w.p.1 to 0 uniformly in $\boldsymbol{\xi}$. Thus, $\varepsilon(\boldsymbol{\xi})$ converges w.p.1 to 0 uniformly in $\boldsymbol{\xi}$.

P.S: If we have $\begin{cases} f_n(\mathbf{x},\boldsymbol{\alpha}) \xrightarrow[uniformly\ in\ \boldsymbol{\alpha}]{P} f(\boldsymbol{\alpha}), \\ Arg\min_{\boldsymbol{\alpha}} f(x) = \boldsymbol{\theta} \end{cases}$ then $\boldsymbol{\theta}_n = Arg\min_{\boldsymbol{\alpha}} f_n(\mathbf{x},\boldsymbol{\alpha}) \xrightarrow{P} \boldsymbol{\theta}$ [Van der Vaart, 1998].

Theorem 1:

Under hypothesizes H1 to H4, the MLE is consistent.

Proof:

From the above lemma, it is sufficient to show that function

$$f(\boldsymbol{\xi}) = \ln|\mathbf{Q}| + tr(\mathbf{Q}^{-1}\mathbf{Q}_0) + tr\left(\mathbf{Q}^{-1}(\mathbf{H}-\mathbf{H}_0)\,\mathbf{R}_{xx}(\mathbf{H}-\mathbf{H}_0)^H\right) \quad (7)$$

has a global minimum for $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ and $\boldsymbol{\mu} = \boldsymbol{\mu}_0$. It is well known that the term $\log|\mathbf{Q}| + tr(\mathbf{Q}^{-1}\mathbf{Q}_0)$ in (7) is minimal iff $\mathbf{Q} = \mathbf{Q}_0$, that is when $\boldsymbol{\mu} = \boldsymbol{\mu}_0$.

Moreover $tr\left(\mathbf{Q}^{-1}(\mathbf{H}-\mathbf{H}_0)\,\mathbf{R}_{xx}(\mathbf{H}-\mathbf{H}_0)^H\right)$ in (7) is positive, and from H4 it admits its minimal value 0 iff $\boldsymbol{\theta} = \boldsymbol{\theta}_0$.

P.S: If two matrices **A** and **B** are positive, then *tr*(AB) is positive [Van der Vaart, 1998].

2.3 Efficiency:

We demonstrate that the ML method is asymptotically efficient for the model (1) of particular observations (3.a).

2.3.1 Fisher Information Matrix:

The Fisher Information Matrix (FIM) for estimating $\xi = \left[\boldsymbol{\theta}^T, \boldsymbol{\mu}^T\right]^T$ from N snapshots is known and easily shown to be given by:

$$\mathbf{F}_N = \begin{bmatrix} \mathbf{F}_{\theta_N} & \mathbf{0} \\ \mathbf{0} & \mathbf{F}_{\mu_N} \end{bmatrix}, \tag{8}$$

where:

$$F_{\theta_N}(i,j) = 2\ \Re\, tr\left(\left(\frac{\partial \mathbf{H}}{\partial \theta_i} \right)_{\boldsymbol{\theta}_0}^H \mathbf{Q}_0^{-1} \left(\frac{\partial \mathbf{H}}{\partial \theta_j} \right)_{\boldsymbol{\theta}_0} \mathbf{X}\mathbf{X}^H \right),$$

for $1 \le i, j \le P_\theta$; (9)

and

$$F_{\mu_N}(i,j) = N\ tr\left(\mathbf{Q}_0^{-1} \left(\frac{\partial \mathbf{Q}}{\partial \mu_i} \right)_{\boldsymbol{\mu}_0} \mathbf{Q}_0^{-1} \left(\frac{\partial \mathbf{Q}}{\partial \mu_j} \right)_{\boldsymbol{\mu}_0} \right), \text{ for}$$

$1 \le i, j \le P_\mu$, (10)

Calculations of the elements of FIM are given in appendix A.

2.3.2 Efficiency:

Let us make the following hypothesis:

H5: H and **Q** are twice derivable with continuous derivatives at $\boldsymbol{\theta}_0$ and $\boldsymbol{\mu}_0$.

Theorem 2: Under hypothesizes H1 to H5 the MLE $\hat{\boldsymbol{\xi}}$ is asymptotically Gaussian and efficient in the following sense:

$\sqrt{N}\left(\hat{\boldsymbol{\xi}} - \boldsymbol{\xi}\right)$ is asymptotically Gaussian, zero mean with covariance matrix $\mathbf{V}^{-1}$ where

$$\mathbf{V} = \lim_{N \longrightarrow \infty} \frac{1}{N} \mathbf{F}_N . \tag{11}$$

Proof is given in appendix B.

Finally, lemmas 2, 4, and the expression (B.5) demonstrate theorem 2.

3. CONCLUSION

In this paper, we have demonstrated that the ML method is consistent and asymptotically efficient for a general model of successive observations **not distributed according the same law of probability**, and we have calculated the corresponding FIM.

We are working on the calculations of the CRB and the application (simulations will be developed) of these results to the Mean Square Error (MSE) of different models of the channel **H**.

REFERENCES

Cramer, H., 1949. Mathematical Methods of Statistics, *Princeton, NJ: Princeton Univ. Press*.

Kassem, H. and P. Forster, 2003. Demodulation with a Multichannel Estimation using a Reduced Rank Constraint. *Kluwer Academic Publishers, Journals Editorial Office, Analog Integrated Circuits and Signal Processing*, 34: 155-162.

Scharf, L.L., 1990. Statistical Signal Processing, Detection, Estimation, and Time Series Analysis. *Addison-Wesly Publishing Company*, 1990, ISBN 0-201-19038-9.

Van der Vaart, A.W., 1998. Asymptotic Statistics, *Combridge University Press* 1998, ISBN 0-521-49603-9.

Appendix A: Calculations of the Elements of FIM:

Unless necessary, the dependence of **s** on $\boldsymbol{\theta}$ will be omitted, that is $\mathbf{s}(t) = \mathbf{s}(t, \boldsymbol{\theta})$

The probability density function of the observation **y**(*t*) can be written as:

$$f(\mathbf{y}(t)) = \frac{1}{\pi^M |\mathbf{Q}|} \exp\left[-(\mathbf{y}(t) - \mathbf{s}(t))^H \mathbf{Q}^{-1} (\mathbf{y}(t) - \mathbf{s}(t)) \right],$$

which conducts for the all N independent observations to:

$$f(\mathbf{y}(1) \cdots \mathbf{y}(N)) = \prod_{t=1}^{N} f(\mathbf{y}(t)) = \frac{1}{\pi^{MN} |\mathbf{Q}|^N} \exp\left[-\sum_{t=1}^{N} (\mathbf{y}(t) - \mathbf{s}(t))^H \mathbf{Q}^{-1} (\mathbf{y}(t) - \mathbf{s}(t)) \right].$$

The corresponding log-likelihood is:

$$L = \ln\left(f(\mathbf{y}(1) \cdots \mathbf{y}(N)) \right) = -MN \ln(\pi) - N \ln|\mathbf{Q}| - \sum_{t=1}^{N} (\mathbf{y}(t) - \mathbf{s}(t))^H \mathbf{Q}^{-1} (\mathbf{y}(t) - \mathbf{s}(t)). \tag{A.1}$$

Partial derivatives:

The partial derivative of L with respect to parameters characterizing **s**(*t*) is:

$$\frac{\partial L}{\partial \theta_p} = -\sum_t \frac{\partial}{\partial \theta_p}\left[(\mathbf{y}(t)-\mathbf{s}(t))^H \mathbf{Q}^{-1}(\mathbf{y}(t)-\mathbf{s}(t))\right]$$

$$= -\sum_t \frac{\partial}{\partial \theta_p}\left[\mathbf{y}(t)^H \mathbf{Q}^{-1}\mathbf{y}(t) - \mathbf{s}(t)^H \mathbf{Q}^{-1}\mathbf{y}(t) - \mathbf{y}(t)^H \mathbf{Q}^{-1}\mathbf{s}(t) + \mathbf{s}(t)^H \mathbf{Q}^{-1}\mathbf{s}(t)\right]$$

$$= \sum_t \mathbf{y}(t)^H \mathbf{Q}^{-1}\frac{\partial s(t)}{\partial \theta_p} + \sum_t \frac{\partial s(t)^H}{\partial \theta_p}\mathbf{Q}^{-1}\mathbf{y}(t) - \sum_t \frac{\partial s(t)^H}{\partial \theta_p}\mathbf{Q}^{-1}s(t) - \sum_t s(t)^H \mathbf{Q}^{-1}\frac{\partial s(t)}{\partial \theta_p} \quad (A.2.a)$$

$$= 2\Re\left\{\sum_t \mathbf{y}(t)^H \mathbf{Q}^{-1}\frac{\partial s(t)}{\partial \theta_p} - \sum_t s(t)^H \mathbf{Q}^{-1}\frac{\partial s(t)}{\partial \theta_p}\right\} (A.2.b)$$

$$= 2\Re\left\{\sum_t (\mathbf{y}(t)-s(t))^H \mathbf{Q}^{-1}\frac{\partial s(t)}{\partial \theta_p}\right\} \quad (A.2.c)$$

Utilizing the identities of matrices [Scharf, 1990]:

- $\frac{\partial}{\partial \mu_k}\ln|\mathbf{Q}| = tr\left(\frac{\partial \mathbf{Q}}{\partial \mu_k}\mathbf{Q}^{-1}\right) = tr\left(\mathbf{Q}^{-1}\frac{\partial \mathbf{Q}}{\partial \mu_k}\right)$;
- $\frac{\partial \mathbf{Q}^{-1}}{\partial \mu_k} = -\mathbf{Q}^{-1}\frac{\partial \mathbf{Q}}{\partial \mu_k}\mathbf{Q}^{-1}$.

Therefore, the expression of partial derivatives of L with respect to the parameters μ_k describing the covariance $\mathbf{Q}$ of the noise:

$$\frac{\partial L}{\partial \mu_k} = -N\,tr\left(\mathbf{Q}^{-1}\frac{\partial \mathbf{Q}}{\partial \mu_k}\right) + \sum_t (\mathbf{y}(t)-\mathbf{s}(t))^H \mathbf{Q}^{-1}\frac{\partial \mathbf{Q}}{\partial \mu_k}\mathbf{Q}^{-1}(\mathbf{y}(t)-\mathbf{s}(t)), \quad (A.3)$$

Calculation of the FIM:

The FIM of the unknown parameters $\boldsymbol{\theta}$ and $\boldsymbol{\mu}$ is defined by:

$$\mathbf{F} = E\begin{bmatrix} \frac{\partial L}{\partial \theta_1}\frac{\partial L}{\partial \theta_1} & \cdots & \frac{\partial L}{\partial \theta_1}\frac{\partial L}{\partial \theta_{p_\theta}} & \frac{\partial L}{\partial \theta_1}\frac{\partial L}{\partial \mu_1} & \cdots & \frac{\partial L}{\partial \theta_1}\frac{\partial L}{\partial \mu_{P_k}} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \frac{\partial L}{\partial \theta_{p_\theta}}\frac{\partial L}{\partial \theta_1} & \cdots & \frac{\partial L}{\partial \theta_{p_\theta}}\frac{\partial L}{\partial \theta_{p_\theta}} & \frac{\partial L}{\partial \theta_{P_\theta}}\frac{\partial L}{\partial \mu_1} & \cdots & \frac{\partial L}{\partial \theta_{p_\theta}}\frac{\partial L}{\partial \mu_{P_k}} \\ \frac{\partial L}{\partial \mu_1}\frac{\partial L}{\partial \theta_1} & \cdots & \frac{\partial L}{\partial \mu_1}\frac{\partial L}{\partial \theta_{p_\theta}} & \frac{\partial L}{\partial \mu_1}\frac{\partial L}{\partial \mu_1} & \cdots & \frac{\partial L}{\partial \mu_1}\frac{\partial L}{\partial \mu_{P_k}} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \frac{\partial L}{\partial \mu_{P_k}}\frac{\partial L}{\partial \theta_1} & \cdots & \frac{\partial L}{\partial \mu_{P_k}}\frac{\partial L}{\partial \theta_{p_\theta}} & \frac{\partial L}{\partial \mu_{P_k}}\frac{\partial L}{\partial \mu_1} & \cdots & \frac{\partial L}{\partial \mu_{P_k}}\frac{\partial L}{\partial \mu_{P_k}} \end{bmatrix} \quad (A.4)$$

Hence, to calculate **F**, we will use the following equalities:

$$E\left[\frac{\partial L}{\partial \theta_i}\frac{\partial L}{\partial \theta_j}\right] = -E\left[\frac{\partial^2 L}{\partial \theta_i \partial \theta_j}\right]$$

and

$$E\left[\frac{\partial L}{\partial \mu_i}\frac{\partial L}{\partial \mu_j}\right] = -E\left[\frac{\partial^2 L}{\partial \mu_i \partial \mu_j}\right].$$

Then we have to calculate the different terms of F.

- **Calculation of** $E\left[\frac{\partial^2 L}{\partial \theta_i \partial \theta_j}\right]$

Starting from (A.2-a), we obtain:

$$\begin{aligned}\frac{\partial^2 L}{\partial\theta_i\partial\theta_j}&=\left(\sum_t \mathbf{y}(t)^H\mathbf{Q}^{-1}\frac{\partial^2\mathbf{s}(t)}{\partial\theta_i\partial\theta_j}\right)+\left(\sum_t \mathbf{y}(t)^H\mathbf{Q}^{-1}\frac{\partial^2\mathbf{s}(t)}{\partial\theta_i\partial\theta_j}\right)^H\\&+\left(-\sum_t\frac{\partial^2\mathbf{s}(t)^H}{\partial\theta_i\partial\theta_j}\mathbf{Q}^{-1}\mathbf{s}(t)-\sum_t\frac{\partial\ \mathbf{s}(t)^H}{\partial\theta_i}\mathbf{Q}^{-1}\frac{\partial\ \mathbf{s}(t)}{\partial\theta_j}\right)\\&+\left(-\sum_t\frac{\partial^2\mathbf{s}(t)^H}{\partial\theta_i\partial\theta_j}\mathbf{Q}^{-1}\mathbf{s}(t)-\sum_t\frac{\partial\ \mathbf{s}(t)^H}{\partial\theta_i}\mathbf{Q}^{-1}\frac{\partial\ \mathbf{s}(t)}{\partial\theta_j}\right)^H\end{aligned}\tag{A.5}$$

Taking the expectation of the precedent expression, we obtain:

$$\begin{aligned}E\left[\frac{\partial^2 L}{\partial\theta_i\partial\theta_j}\right]&=\sum_t\mathbf{s}(t)^H\mathbf{Q}^{-1}\frac{\partial^2\mathbf{s}(t)}{\partial\theta_i\partial\theta_j}+\sum_t\frac{\partial^2\mathbf{s}(t)^H}{\partial\theta_i\theta_j}\mathbf{Q}^{-1}\mathbf{s}(t)\\&-\sum_t\frac{\partial^2\mathbf{s}(t)^H}{\partial\theta_i\theta_j}\mathbf{Q}^{-1}\mathbf{s}(t)-\sum_t\frac{\partial\mathbf{s}(t)^H}{\partial\theta_i}\mathbf{Q}^{-1}\frac{\partial\mathbf{s}(t)}{\partial\theta_j}\\&-\sum_t\mathbf{s}(t)^H\mathbf{Q}^{-1}\frac{\partial^2\mathbf{s}(t)}{\partial\theta_i\partial\theta_j}-\sum_t\frac{\partial\mathbf{s}(t)^H}{\partial\theta_j}\mathbf{Q}^{-1}\frac{\partial\mathbf{s}(t)}{\partial\theta_i}\\&=-2\Re\left[\sum_t\frac{\partial\mathbf{s}(t)^H}{\partial\theta_i}\mathbf{Q}^{-1}\frac{\partial\ \mathbf{s}(t)}{\partial\theta_j}\right]\\&=-2\left[\sum_t\frac{\partial\mathbf{s}(t)^H}{\partial\theta_i}\mathbf{Q}^{-1}\frac{\partial\mathbf{s}(t)}{\partial\theta_j}\right]\end{aligned}\tag{A.6}$$

- **Calculation of** $E\left[\frac{\partial^2 L}{\partial\mu_i\partial\mu_j}\right]$

Starting from (A.3), we have:

$$\begin{aligned}\frac{\partial L}{\partial\mu_k}&=-Ntr\left(\mathbf{Q}^{-1}\frac{\partial\mathbf{Q}}{\partial\mu_k}\right)+\sum_t\left(\mathbf{y}(t)-\mathbf{s}(t)\right)^H\mathbf{Q}^{-1}\frac{\partial\mathbf{Q}}{\partial\mu_k}\mathbf{Q}^{-1}\left(\mathbf{y}(t)-\mathbf{s}(t)\right)\\&=-Ntr\left(\mathbf{Q}^{-1}\frac{\partial\mathbf{Q}}{\partial\mu_k}\right)+\sum_t\mathbf{n}(t)^H\mathbf{Q}^{-1}\frac{\partial\mathbf{Q}}{\partial\mu_k}\mathbf{Q}^{-1}\mathbf{n}(t)\end{aligned}\tag{A.7}$$

thus,

$$\begin{aligned}\frac{\partial^2 L}{\partial\mu_i\partial\mu_j}&=-Ntr\left(\mathbf{Q}^{-1}\frac{\partial^2\mathbf{Q}}{\partial\mu_i\partial\mu_j}\right)+Ntr\left(\mathbf{Q}^{-1}\frac{\partial\mathbf{Q}}{\partial\mu_j}\mathbf{Q}^{-1}\frac{\partial\mathbf{Q}}{\partial\mu_i}\right)+\sum_t\mathbf{n}(t)^H\mathbf{Q}^{-1}\frac{\partial^2\mathbf{Q}}{\partial\mu_i\partial\mu_j}\mathbf{Q}^{-1}\mathbf{n}(t)\\&-\sum_t\mathbf{n}(t)^H\mathbf{Q}^{-1}\frac{\partial\mathbf{Q}}{\partial\mu_j}\mathbf{Q}^{-1}\frac{\partial\mathbf{Q}}{\partial\mu_i}\mathbf{Q}^{-1}\mathbf{n}(t)-\sum_t\mathbf{n}(t)^H\mathbf{Q}^{-1}\frac{\partial\mathbf{Q}}{\partial\mu_i}\mathbf{Q}^{-1}\frac{\partial\mathbf{Q}}{\partial\mu_j}\mathbf{Q}^{-1}\mathbf{n}(t)\\&=-Ntr\left(\mathbf{Q}^{-1}\frac{\partial^2\mathbf{Q}}{\partial\mu_i\partial\mu_j}\right)+Ntr\left(\mathbf{Q}^{-1}\frac{\partial\mathbf{Q}}{\partial\mu_j}\mathbf{Q}^{-1}\frac{\partial\mathbf{Q}}{\partial\mu_i}\right)\\&+tr\left[\mathbf{Q}^{-1}\frac{\partial^2\mathbf{Q}}{\partial\mu_i\partial\mu_j}\mathbf{Q}^{-1}\sum_t\mathbf{n}(t)\mathbf{n}(t)^H\right]-tr\left[\mathbf{Q}^{-1}\frac{\partial\mathbf{Q}}{\partial\mu_j}\mathbf{Q}^{-1}\frac{\partial\mathbf{Q}}{\partial\mu_i}\mathbf{Q}^{-1}\sum_t\mathbf{n}(t)\mathbf{n}(t)^H\right]\\&-tr\left[\mathbf{Q}^{-1}\frac{\partial\mathbf{Q}}{\partial\mu_i}\mathbf{Q}^{-1}\frac{\partial\mathbf{Q}}{\partial\mu_j}\mathbf{Q}^{-1}\sum_t\mathbf{n}(t)\mathbf{n}(t)^H\right]\end{aligned}\tag{A.8}$$

Taking the expectation of the precedent expression, we obtain

$$\begin{aligned}E\left[\frac{\partial^2 L}{\partial\mu_i\partial\mu_j}\right]&=-Ntr\left(\mathbf{Q}^{-1}\frac{\partial^2\mathbf{Q}}{\partial\mu_i\partial\mu_j}\right)+Ntr\left(\mathbf{Q}^{-1}\frac{\partial\mathbf{Q}}{\partial\mu_i}\mathbf{Q}^{-1}\frac{\partial\mathbf{Q}}{\partial\mu_j}\right)+Ntr\left(\mathbf{Q}^{-1}\frac{\partial^2\mathbf{Q}}{\partial\mu_i\partial\mu_j}\right)\\&-Ntr\left(\mathbf{Q}^{-1}\frac{\partial\mathbf{Q}}{\partial\mu_i}\mathbf{Q}^{-1}\frac{\partial\mathbf{Q}}{\partial\mu_j}\right)-Ntr\left(\mathbf{Q}^{-1}\frac{\partial\mathbf{Q}}{\partial\mu_i}\mathbf{Q}^{-1}\frac{\partial\mathbf{Q}}{\partial\mu_j}\right)\\&=-Ntr\left(\mathbf{Q}^{-1}\frac{\partial\mathbf{Q}}{\partial\mu_i}\mathbf{Q}^{-1}\frac{\partial\mathbf{Q}}{\partial\mu_j}\right)\end{aligned}\tag{A.9}$$

- **Calculation of** $E\left[\frac{\partial^2 L}{\partial\theta_i\partial\mu_j}\right]$

Starting from (A.2-c), we have:

$$\frac{\partial L}{\partial\theta_p}=2\Re\left\{\sum_t\left(\mathbf{y}(t)-\mathbf{s}(t)\right)^H\mathbf{Q}^{-1}\frac{\partial\mathbf{s}(t)}{\partial\theta_p}\right\},$$

thus

$$\frac{\partial^2 L}{\partial\theta_i\partial\mu_j}=2\Re\left\{\sum_t\left(\mathbf{y}(t)-\mathbf{s}(t)\right)^H\frac{\partial\mathbf{Q}^{-1}}{\partial\mu_j}\frac{\partial\mathbf{s}(t)}{\partial\theta_i}\right\},$$

which conducts to

$$E\left[\frac{\partial^2 L}{\partial\theta_i\partial\mu_j}\right]=2\Re\left\{\sum_t E\left[\left(\mathbf{y}(t)-\mathbf{s}(t)\right)^H\right]\frac{\partial\mathbf{Q}^{-1}}{\partial\mu_j}\frac{\partial\mathbf{s}(t)}{\partial\theta_i}\right\}=0\cdot$$

The FIM can be written using the terms calculated previously:

$$\mathbf{F}=\begin{bmatrix}\mathbf{F}_\theta & \mathbf{0}\\ \mathbf{0} & \mathbf{F}_\mu\end{bmatrix}, \tag{8}$$

where:

$$F_\theta(i,j)=2\,\Re\sum_{t=1}^{N}\left(\frac{\partial\mathbf{s}(t,\boldsymbol{\theta})}{\partial\theta_i}\right)^H\mathbf{Q}^{-1}\left(\frac{\partial\mathbf{s}(t,\boldsymbol{\theta})}{\partial\theta_j}\right),\ \text{for } 1\le i,j\le P_\theta, \tag{9}$$

$$F_\mu(i,j)=N\,tr\left(\mathbf{Q}^{-1}\left(\frac{\partial\mathbf{Q}}{\partial\mu_i}\right)\mathbf{Q}^{-1}\left(\frac{\partial\mathbf{Q}}{\partial\mu_j}\right)\right),\ \text{for } 1\le i,j\le P_\mu\cdot \tag{10}$$

Appendix B: Proof of Theorem 2:

First note that $\lim_{N\longrightarrow\infty}\frac{1}{N}\mathbf{F}_N$ which is involved in the above theorem is well defined thanks to H3. Indeed, we have from (9):

$$\frac{1}{N}F_{\theta_N}(i,j)=\frac{2}{N}\Re tr\left(\left(\frac{\partial\mathbf{H}}{\partial\theta_i}\right)_{\boldsymbol{\theta}_0}^H\mathbf{Q}_0^{-1}\left(\frac{\partial\mathbf{H}}{\partial\theta_j}\right)_{\boldsymbol{\theta}_0}\mathbf{X}\mathbf{X}^H\right)\xrightarrow{N\to\infty}2\Re tr\left(\left(\frac{\partial\mathbf{H}}{\partial\theta_i}\right)_{\boldsymbol{\theta}_0}^H\mathbf{Q}_0^{-1}\left(\frac{\partial\mathbf{H}}{\partial\theta_j}\right)_{\boldsymbol{\theta}_0}\mathbf{R}_{xx}\right) \tag{B.1}$$

and obviously:

$$\frac{1}{N}F_{\mu_N}(i,j)=tr\left(\mathbf{Q}_0^{-1}\left(\frac{\partial\mathbf{Q}}{\partial\mu_i}\right)_{\boldsymbol{\mu}_0}\mathbf{Q}_0^{-1}\left(\frac{\partial\mathbf{Q}}{\partial\mu_j}\right)_{\boldsymbol{\mu}_0}\right). \tag{B.2}$$

Thus $\lim_{N\longrightarrow\infty}\frac{1}{N}\mathbf{F}_N$ is well defined.

Next, a first order Taylor expansion of the gradient of the ML criterion (3) yields:

$$\left(\frac{\partial L}{\partial\xi_i}\right)_{\hat{\xi}}=\left(\frac{\partial L}{\partial\xi_i}\right)_{\xi_0}+\sum_{j=1}^{P_\theta+P_\mu}\left(\frac{\partial^2 L}{\partial\xi_i\partial\xi_j}\right)_{\xi'}\left(\hat{\xi}_j-\xi_{j_0}\right)=0, \tag{B.3}$$

where

$$\xi'=\gamma\ \ \xi_0+(1-\gamma)\ \ \hat{\xi}\ \text{ with }\gamma\in[0\ \ 1]. \tag{B.4}$$
$$\overset{\Delta}{=}\begin{pmatrix}\boldsymbol{\theta}'\\ \boldsymbol{\mu}'\end{pmatrix}$$

Thus,

$$\hat{\xi}-\xi_0=-\mathbf{H}(\xi')^{-1}\mathbf{G}, \tag{B.5}$$

where $\mathbf{G}=(G_i)$ and $\mathbf{H}(\xi')=\mathbf{H}_{ij}(\xi')$ denote respectively the gradient at ξ_0 and the Hessian at ξ':

- $G_i=\left(\frac{\partial L}{\partial\xi_i}\right)_{\xi_0}$; (B.6)
- $H_{ij}(\xi')=\left(\frac{\partial^2 L}{\partial\xi_i\partial\xi_j}\right)_{\xi'}$. (B.7)

The statistical properties of the gradient **G** and Hessian $\mathbf{H}(\xi^{'})$ are summarized in lemmas 2 and 4 here below.

Lemma 2: $\sqrt{N}\mathbf{G}$, where **G** is defined by (B.6), is asymptotically Gaussian with zero-mean and covariance matrix $\lim_{N\longrightarrow\infty}\frac{1}{N}\mathbf{F}_N$.

Proof: The components of the gradient with respect to μ_k are given by:

$$\left(\frac{\partial L}{\partial \mu_k}\right)_{\xi_0} = tr\left(\left(\frac{\partial \mathbf{Q}}{\partial \mu_k}\right)_{\mu_0}\mathbf{Q}_0^{-1}\right) - N^{-1}tr\left(\mathbf{Q}_0^{-1}\left(\frac{\partial \mathbf{Q}}{\partial \mu_k}\right)_{\mu_0}\mathbf{Q}_0^{-1}\,\mathbf{N}\mathbf{N}^H\right) \qquad (B.8)$$

$$= -tr\left(\mathbf{Q}_0^{-1}\left(\frac{\partial \mathbf{Q}}{\partial \mu_k}\right)_{\mu_0}\mathbf{Q}_0^{-1}\left(N^{-1}\mathbf{N}\mathbf{N}^H - \mathbf{Q}_0\right)\right)$$

Since $E[N^{-1}\mathbf{B}\mathbf{B}^H] = \mathbf{Q}_0$, we have: $E\left[\left(\frac{\partial L}{\partial \mu_k}\right)_{\xi_0}\right] = 0$.

Similarly

$$\left(\frac{\partial L}{\partial \theta_p}\right)_{\xi_0} = -N^{-1}tr\left[\mathbf{Q}_0^{-1}\left(\left(\frac{\partial \mathbf{H}}{\partial \theta_p}\right)_{\theta_0}\mathbf{X}\mathbf{N}^H + \mathbf{N}\mathbf{X}^H\left(\frac{\partial \mathbf{H}^H}{\partial \theta_p}\right)_{\theta_0}\right)\right], \qquad (B.9)$$

which implies, since **B** is zero-mean: $E\left[\left(\frac{\partial L}{\partial \theta_p}\right)_{\xi_0}\right] = 0$.

Thus, the gradient **G** is zero-mean. Furthermore, we see from expression (B.9) that $\left(\frac{\partial L}{\partial \theta_p}\right)_{\xi_0}$ is Gaussian, while the central limit theorem applied to $N^{-1}\mathbf{N}\mathbf{N}^H = N^{-1}\sum_{t=1}^{N}\mathbf{n}(t)\mathbf{n}(t)^H$ in (B.8) implies that $\sqrt{N}\,E\left[\left(\frac{\partial L}{\partial \mu_k}\right)_{\xi_0}\right]$ is asymptotically Gaussian: therefore $\sqrt{N}\mathbf{G}$ is asymptotically Gaussian. Finally, a simple computation shows that the covariance of the elements $\sqrt{N}\mathbf{G}$ is given by $\frac{1}{N}\mathbf{F}_N$ which completes the proof of lemma 2.

Computation of the Covariance of the Elements:

The Gradient **G** is defined by (B.6): $G_i = \left(\frac{\partial L}{\partial \xi_i}\right)_{\xi_0}$

- **Analysis of** $N.E\left[\frac{\partial L}{\partial \mu_i}\frac{\partial L}{\partial \mu_j}\right]$

Starting from (B.8), we have: $\left(\frac{\partial L}{\partial \mu_k}\right)_{\xi_0} = -tr\left(\mathbf{Q}_0^{-1}\left(\frac{\partial \mathbf{Q}}{\partial \mu_k}\right)_{\mu_0}\mathbf{Q}_0^{-1}\left(N^{-1}\mathbf{N}\mathbf{N}^H - \mathbf{Q}_0\right)\right)$,

and

$$N.E\left[\frac{\partial L}{\partial \mu_i}\frac{\partial L}{\partial \mu_j}\right] = N.E\left[tr\left\{\mathbf{Q}_0^{-1}\left(\frac{\partial \mathbf{Q}}{\partial \mu_i}\right)_{\mu_0}\mathbf{Q}_0^{-1}\left(\frac{1}{N}\mathbf{N}\mathbf{N}^H - \mathbf{Q}_0\right)\right\}tr\left\{\mathbf{Q}_0^{-1}\left(\frac{\partial \mathbf{Q}}{\partial \mu_j}\right)_{\mu_0}\mathbf{Q}_0^{-1}\left(\frac{1}{N}\mathbf{N}\mathbf{N}^H - \mathbf{Q}_0\right)\right\}\right]$$

where **N** is circular complex Gaussian, then [Van der Vaart, 1998]:

$$E\left\{tr\left[\left(\hat{\mathbf{Q}} - \mathbf{Q}_0\right)\mathbf{A}\right]tr\left[\left(\hat{\mathbf{Q}} - \mathbf{Q}_0\right)\mathbf{B}\right]\right\} = \frac{1}{N}tr\left[\mathbf{Q}_0\ \mathbf{A}\ \mathbf{Q}_0\mathbf{B}\right],$$

We deduce:

$$N.E\left[\frac{\partial L}{\partial \mu_i}\frac{\partial L}{\partial \mu_j}\right] = tr\left\{\mathbf{Q}_0^{-1}\left(\frac{\partial \mathbf{Q}}{\partial \mu_i}\right)_{\mu_0}\mathbf{Q}_0^{-1}\mathbf{Q}_0\,\mathbf{Q}_0^{-1}\left(\frac{\partial \mathbf{Q}}{\partial \mu_j}\right)_{\mu_0}\mathbf{Q}_0^{-1}\mathbf{Q}_0\right\}$$

$$= tr\left\{\mathbf{Q}_0^{-1}\left(\frac{\partial \mathbf{Q}}{\partial \mu_i}\right)_{\mu_0}\mathbf{Q}_0^{-1}\left(\frac{\partial \mathbf{Q}}{\partial \mu_j}\right)_{\mu_0}\right\}$$

$$= \frac{1}{N}\mathbf{F}_{\mu}$$

- **Analysis of** $N.E\left[\frac{\partial L}{\partial \mu_i}\frac{\partial L}{\partial \theta_j}\right]$

$N.E\left[\frac{\partial L}{\partial \mu_i}\frac{\partial L}{\partial \theta_j}\right]=\mathbf{0}$, Moments of odd order of **N**

- **Analysis of** $N.E\left[\frac{\partial L}{\partial \theta_i}\frac{\partial L}{\partial \theta_j}\right]$

Starting from (B.9), we have:

$$\left(\frac{\partial L}{\partial \theta_p}\right)_{\xi_0}=-N^{-1}\left\{\sum_{t=1}^{N}\mathbf{n}(t)^H\mathbf{Q}_0^{-1}\left(\frac{\partial \mathbf{s}(t)}{\partial \theta_p}\right)_{\xi_0}+\sum_t\left(\frac{\partial \mathbf{s}(t)^H}{\partial \theta_p}\right)_{\xi_0}\mathbf{Q}_0^{-1}\mathbf{n}(t)\right\},$$

and

$$\begin{aligned}N.E\left[\frac{\partial L}{\partial \theta_i}\frac{\partial L}{\partial \theta_j}\right]&=N^{-1}\left(\sum_t E\left\{\mathbf{n}(t)^H\mathbf{Q}_0^{-1}\left(\frac{\partial \mathbf{s}(t)}{\partial \theta_i}\right)_{\xi_0}\left(\frac{\partial \mathbf{s}(t)^H}{\partial \theta_j}\right)_{\xi_0}\mathbf{Q}_0^{-1}\mathbf{n}(t)\right\}\right)\\&+N^{-1}\left(\sum_t E\left\{\mathbf{n}(t)^H\mathbf{Q}_0^{-1}\left(\frac{\partial \mathbf{s}(t)}{\partial \theta_j}\right)_{\xi_0}\left(\frac{\partial \mathbf{s}(t)^H}{\partial \theta_i}\right)_{\xi_0}\mathbf{Q}_0^{-1}\mathbf{n}(t)\right\}\right)\\&+N^{-1}\left(\sum_t E\left\{\mathbf{n}(t)^H\mathbf{Q}_0^{-1}\left(\frac{\partial \mathbf{s}(t)}{\partial \theta_j}\right)_{\xi_0}\mathbf{n}(t)^H\mathbf{Q}_0^{-1}\left(\frac{\partial \mathbf{s}(t)}{\partial \theta_i}\right)_{\xi_0}\right\}\right)\\&+N^{-1}\left(\sum_t E\left\{\left(\frac{\partial \mathbf{s}(t)^H}{\partial \theta_j}\right)_{\xi_0}\mathbf{Q}_0^{-1}\mathbf{n}(t)\left(\frac{\partial \mathbf{s}(t)^H}{\partial \theta_i}\right)_{\xi_0}\mathbf{Q}_0^{-1}\mathbf{n}(t)\right\}\right)\\&=N^{-1}\left(\sum_t\left\{\left(\frac{\partial \mathbf{s}(t)^H}{\partial \theta_j}\right)_{\xi_0}\mathbf{Q}_0^{-1}\left(\frac{\partial \mathbf{s}(t)}{\partial \theta_i}\right)_{\xi_0}\right\}+\left(\sum_t\left\{\left(\frac{\partial \mathbf{s}(t)^H}{\partial \theta_j}\right)_{\xi_0}\mathbf{Q}_0^{-1}\left(\frac{\partial \mathbf{s}(t)}{\partial \theta_i}\right)_{\xi_0}\right\}\right)^H\right)\\&=2\,\Re\,N^{-1}\left(\sum_t\left\{\left(\frac{\partial \mathbf{s}(t)^H}{\partial \theta_i}\right)_{\xi_0}\mathbf{Q}_0^{-1}\left(\frac{\partial \mathbf{s}(t)}{\partial \theta_j}\right)_{\xi_0}\right\}\right)\\&=\frac{1}{N}\mathbf{F}_\theta\end{aligned}$$

We deduce that the covariance of $\sqrt{N}\,\mathbf{G}$ is given by $\frac{1}{N}\mathbf{F}_N$.

Lemma 3*:* $N^{-1}\mathbf{X}\mathbf{N}^H\xrightarrow{P}\mathbf{0}$ as *N* tends to infinity.

Proof*:* The elements of $N^{-1}\mathbf{X}\mathbf{N}^H$ are zero-mean complex Gaussian random variables as those of **N**. Let us demonstrate that the sum of the variances of their real and imaginary parts $E\left[\left\|N^{-1}\mathbf{X}\mathbf{N}^H\right\|_F^2\right]$ tends to zero as $N\longrightarrow\infty$ which implies lemma 3. Indeed, it is easy to show that:

$$\begin{aligned}E\left[\left\|N^{-1}\ \mathbf{X}\ \mathbf{N}^H\right\|_F^2\right]&=E\left[tr\left(N^{-2}\ \mathbf{X}\ \mathbf{N}^H\mathbf{N}\ \mathbf{X}^H\right)\right]\\&\le N^{-1}tr\ \mathbf{Q}_0\ tr\left(N^{-1}\ \mathbf{X}\ \mathbf{X}^H\right)\end{aligned}$$

which tends to zero as $N\longrightarrow\infty$ from H3.

P.S: $\|\mathbf{AB}\|_F\le\|\mathbf{A}\|_F\|\mathbf{B}\|_F$ and $tr(\mathbf{AB})\ne tr(\mathbf{A})tr(\mathbf{B})$.

Corollary*:* If $\boldsymbol{\theta}\xrightarrow{P}\boldsymbol{\theta}_0$ as $N\longrightarrow\infty$, then $N^{-1}\ (\mathbf{Y}-\mathbf{S})\ (\mathbf{Y}-\mathbf{S})^H\xrightarrow{P}\mathbf{Q}_0$.

Proof: Let us split $N^{-1}(\mathbf{Y}-\mathbf{S})(\mathbf{Y}-\mathbf{S})^H$ into:

$$\begin{aligned} N^{-1}(\mathbf{Y}-\mathbf{S})(\mathbf{Y}-\mathbf{S})^H &= N^{-1}(\mathbf{S}_0+\mathbf{N}-\mathbf{S})(\mathbf{S}_0+\mathbf{N}-\mathbf{S})^H \\ &= N^{-1}\mathbf{N}\mathbf{N}^H+(\mathbf{H}-\mathbf{H}_0)\left(N^{-1}\mathbf{X}\mathbf{X}^H\right)(\mathbf{H}-\mathbf{H}_0)^H \\ &\quad -(\mathbf{H}-\mathbf{H}_0)(N^{-1}\mathbf{X}\mathbf{N}^H)-(N^{-1}\mathbf{N}\,\mathbf{X}^H)(\mathbf{H}-\mathbf{H}_0)^H \end{aligned} \tag{B.10}$$

In the above decomposition, we have from the weak law of large numbers.

$$N^{-1}\mathbf{N}\mathbf{N}^H \xrightarrow{P} \mathbf{Q}_0 \tag{B.11}$$

In the second term $(\mathbf{H}-\mathbf{H}_0)\left(N^{-1}\mathbf{X}\mathbf{X}^H\right)(\mathbf{H}-\mathbf{H}_0)^H$, the continuity of $\mathbf{H}$ (which results from H5) implies that

$$\mathbf{H} \xrightarrow{p} \mathbf{H}_0 \text{ as } \boldsymbol{\theta} \xrightarrow{P} \boldsymbol{\theta}_0. \tag{B.12}$$

It follows from H3 that:

$$(\mathbf{H}-\mathbf{H}_0)\left(N^{-1}\mathbf{X}\mathbf{X}^H\right)(\mathbf{H}-\mathbf{H}_0)^H \xrightarrow{P} 0 \text{ as } N \longrightarrow \infty. \tag{B.13}$$

Finally, from lemma 3 and equation (B.12), the last two terms in (B.10) converge w.p.1 to 0.

Lemma 4*:* The Hessian matrix $\mathbf{H}(\boldsymbol{\xi}')$ defined by (B.7) converges w.p.1 to $\mathbf{V}$ (11).

Proof: Let us study the behavior of the three terms $\left(\frac{\partial^2 L}{\partial\mu_i\partial\mu_j}\right)_{\xi'}$, $\left(\frac{\partial^2 L}{\partial\mu_i\partial\theta_j}\right)_{\xi'}$, and $\left(\frac{\partial^2 L}{\partial\theta_i\partial\theta_j}\right)_{\xi'}$ as $N \longrightarrow \infty$.

- **Analysis of** $\left(\frac{\partial^2 L}{\partial\mu_i\partial\mu_j}\right)_{\xi'}$

We have got:

$$\left(\frac{\partial L}{\partial\mu_i}\right)_{\xi'} = tr\left(\mathbf{Q}^{-1}\frac{\partial\mathbf{Q}}{\partial\mu_i}\right)-N^{-1}tr\left(\mathbf{Q}^{-1}\frac{\partial\mathbf{Q}}{\partial\mu_i}\mathbf{Q}^{-1}(\mathbf{Y}-\mathbf{S})(\mathbf{Y}-\mathbf{S})^H\right),$$

From which we deduce:

$$\begin{aligned} \left(\frac{\partial^2 L}{\partial\mu_i\partial\mu_j}\right)_{\xi'} &= -tr\left(\mathbf{Q}^{-1}\frac{\partial\mathbf{Q}}{\partial\mu_i}\mathbf{Q}^{-1}\frac{\partial\mathbf{Q}}{\partial\mu_j}\right)+tr\left(\mathbf{Q}^{-1}\frac{\partial^2\mathbf{Q}}{\partial\mu_i\partial\mu_j}\right) \\ &\quad -N^{-1}tr\left(\mathbf{Q}^{-1}\frac{\partial^2\mathbf{Q}}{\partial\mu_i\partial\mu_j}\mathbf{Q}^{-1}(\mathbf{Y}-\mathbf{S})(\mathbf{Y}-\mathbf{S})^H\right) \\ &\quad +N^{-1}tr\left(\mathbf{Q}^{-1}\frac{\partial\mathbf{Q}}{\partial\mu_j}\mathbf{Q}^{-1}\frac{\partial\mathbf{Q}}{\partial\mu_i}\mathbf{Q}^{-1}(\mathbf{Y}-\mathbf{S})(\mathbf{Y}-\mathbf{S})^H\right) \\ &\quad +N^{-1}tr\left(\mathbf{Q}^{-1}\frac{\partial\mathbf{Q}}{\partial\mu_i}\mathbf{Q}^{-1}\frac{\partial\mathbf{Q}}{\partial\mu_j}\mathbf{Q}^{-1}(\mathbf{Y}-\mathbf{S})(\mathbf{Y}-\mathbf{S})^H\right) \end{aligned} \tag{B.14}$$

The consistence of MLE implies $\boldsymbol{\xi}' \xrightarrow{P} \boldsymbol{\xi}_0$.

where:

$$\boldsymbol{\xi}' = \gamma\boldsymbol{\xi}_0+(1-\gamma)\hat{\boldsymbol{\xi}} \text{ with} \gamma \in [0\ 1]$$

Therefore, in the second member of (B.14) we have:

Australian Journal of Basic and Applied Sciences, 9(11) May 2015, Pages: 14-23

$$\mathbf{Q} \xrightarrow{P} \mathbf{Q}_0 \text{ and } \left(\frac{\partial^2 L}{\partial \mu_i \partial \mu_j}\right)_{\xi'} \xrightarrow{P} \left(\frac{\partial^2 L}{\partial \mu_i \partial \mu_j}\right)_{\xi_0}.$$

Furthermore, according to the corollary of lemma 3, we have in the second member of (B.14):

$$N^{-1}(\mathbf{Y}-\mathbf{S})(\mathbf{Y}-\mathbf{S})^H \xrightarrow{P} \mathbf{Q}_0.$$

Then:

$$\left(\frac{\partial^2 L}{\partial \mu_i \partial \mu_j}\right)_{\xi'} \xrightarrow{P} tr\left(\mathbf{Q}_0^{-1}\left(\frac{\partial \mathbf{Q}}{\partial \mu_i}\right)_{\xi_0}\mathbf{Q}_0^{-1}\left(\frac{\partial \mathbf{Q}}{\partial \mu_j}\right)_{\xi_0}\right). \tag{B.15}$$

- **Analysis of** $\left(\frac{\partial^2 L}{\partial \mu_i \partial \theta_j}\right)_{\xi'}$

$$\begin{aligned}\left(\frac{\partial^2 L}{\partial \mu_i \partial \theta_j}\right)_{\xi'} &= N^{-1}tr\left(\mathbf{Q}^{-1}\frac{\partial \mathbf{Q}}{\partial \mu_i}\mathbf{Q}^{-1}\frac{\partial \mathbf{S}}{\partial \theta_j}(\mathbf{S}_0-\mathbf{S}+\mathbf{N})^H\right)+N^{-1}tr\left(\mathbf{Q}^{-1}\frac{\partial \mathbf{Q}}{\partial \mu_i}\mathbf{Q}^{-1}(\mathbf{S}_0-\mathbf{S}+\mathbf{N})\frac{\partial \mathbf{S}^H}{\partial \theta_j}\right)\\ &= 2\Re\left\{tr\left(\mathbf{Q}^{-1}\frac{\partial \mathbf{Q}}{\partial \mu_i}\mathbf{Q}^{-1}\frac{\partial \mathbf{H}}{\partial \theta_j}\frac{\mathbf{X}\mathbf{X}^H}{N}(\mathbf{H}_0-\mathbf{H})^H\right)\right\}+2\Re\left\{tr\left(\mathbf{Q}^{-1}\frac{\partial \mathbf{Q}}{\partial \mu_i}\mathbf{Q}^{-1}\frac{\partial \mathbf{H}}{\partial \theta_j}\frac{\mathbf{X}\mathbf{N}^H}{N}\right)\right\}\end{aligned} \tag{B.16}$$

We easily verify that the two terms of the second member in (B.16) converge in probability to 0 by utilizing hypothesis H3, lemma 3 and $\xi' \xrightarrow{P} \xi_0$:

$$\left(\frac{\partial^2 L}{\partial \mu_i \partial \theta_j}\right)_{\xi'} \xrightarrow{P} 0. \tag{B.17}$$

- **Analysis of** $\left(\frac{\partial^2 L}{\partial \theta_i \partial \theta_j}\right)_{\xi'}$

$$\begin{aligned}\left(\frac{\partial^2 L}{\partial \theta_i \partial \theta_j}\right)_{\xi'} &= 2N^{-1}\Re tr\left(\frac{\partial \mathbf{S}^H}{\partial \theta_i}\mathbf{Q}^{-1}\frac{\partial \mathbf{S}}{\partial \theta_j}-\mathbf{Q}^{-1}\frac{\partial^2 \mathbf{S}}{\partial \theta_i \partial \theta_j}(\mathbf{S}_0-\mathbf{S}+\mathbf{N})^H\right)\\ &= 2N^{-1}\Re tr\left(\mathbf{X}^H\left(\frac{\partial \mathbf{H}}{\partial \theta_i}\right)^H\mathbf{Q}^{-1}\left(\frac{\partial \mathbf{H}}{\partial \theta_j}\right)\mathbf{X}-\mathbf{Q}^{-1}\frac{\partial^2 \mathbf{H}}{\partial \theta_i \partial \theta_j}\mathbf{X}\left((\mathbf{H}_0-\mathbf{H})\mathbf{X}+\mathbf{N}\right)^H\right)\\ &= 2\ \Re\ tr\left(\frac{\mathbf{X}\mathbf{X}^H}{N}\left(\frac{\partial \mathbf{H}}{\partial \theta_i}\right)^H\mathbf{Q}^{-1}\left(\frac{\partial \mathbf{H}}{\partial \theta_j}\right)\right)-2\ \Re\ tr\left(\mathbf{Q}^{-1}\frac{\partial^2 \mathbf{H}}{\partial \theta_i \partial \theta_j}\left(\frac{\mathbf{X}\mathbf{X}^H}{N}(\mathbf{H}_0-\mathbf{H})^H+\frac{\mathbf{X}\mathbf{N}^H}{N}\right)\right)\end{aligned} \tag{B.18}$$

We easily verify that the two terms of the second member in (B.18) converge in probability to 0 by utilizing hypothesis H3, lemma 3 and $\xi' \xrightarrow{P} \xi_0$:

$$\left(\frac{\partial^2 L}{\partial \theta_i \partial \theta_j}\right)_{\xi'} \xrightarrow{P} 2\Re tr\left(\mathbf{R}_{xx}\left(\frac{\partial \mathbf{H}}{\partial \theta_i}\right)_{\xi_0}^H\mathbf{Q}_0^{-1}\left(\frac{\partial \mathbf{H}}{\partial \theta_j}\right)_{\xi_0}\right). \tag{B.19}$$

Thus, $\mathbf{H}(\nu')$ converges w.p.1 to $\mathbf{V} = \lim_{N\longrightarrow\infty}\frac{1}{N}\mathbf{F}_N$