Consistency and Efficiency of the Maximum Likelihood Method of Independent and Different Gaussian Observations

Seifedine Kadry LIU, Bekaa Campus, Al-Khyara Lebanon Hussam Kassem

Lebanese International University, Faculty of Engineering, Lebanon
P.O. Box 5, Jeb Janeen
Lebanese University, Faculty of Science, Lebanon

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 for similar results in the case of independent Gaussian observations that are not distributed according to the same law. The corresponding Cramer-Rao Bound (CRB) is determined and applied to the calculation of the Mean Square Error (MSE) of different models of the channel. Simulations show that ML Estimator (MLE) is consistent and efficient.

Key words: Maximum Likelihood Estimator, Consistency, Efficiency, Gaussian Observations, Signal Processing.

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 radiomobile 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, either data important or adaptive. 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 and the efficiency of the MLE, section 3 calculates the corresponding Fisher Information Matrix (FIM) and Cramer-Rao Bound (CRB), and gives their applications to the calculation of the Mean Square Error (MSE) of different models of the channel. Simulations are given in section 4.

Consistency and Efficiency:

2.1 Data Model:

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

\[ y(t) = s(t, \theta_0) + n(t) \quad \text{or} \quad t = 1, \cdots, M \] (1)

where:

- \( y(t) \) is the noisy measured signal with \( M \) components;
- \( s(t, \theta) \) is a deterministic information bearing signal that depends on an unknown vector parameter
- \( \theta = [\theta_1 \cdots \theta_P]^T \) whose exact value is \( \theta_0 \)
- \( n(t) \) is a temporally white complex Gaussian, circular, zero-mean noise with covariance matrix \( Q(\mu) \) where

Corresponding Author: Seifedine Kadry, Lebanese International University, Faculty of Engineering, Lebanon; P. O. Box 5, Jeb Janeen
E-mail: hussamkassem@yahoo.com
\[ \mathbf{\mu} = \begin{bmatrix} \mu_1 & \cdots & \mu_p \end{bmatrix}^T \text{ an unknown vector parameter.} \]

Thus, observations \( y(t) \) 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 radiocommunications, we have

\[ s(t, \mathbf{\theta}) = \sum_{k=1}^p \beta_k s(t - \tau_k) a(x_k) \]

where \( d(t) \) is a known signal, and \( \beta_k, \tau_k \) are unknown multipath parameters (angle of arrival, complex amplitude and delay). In the following we set

\[ \mathbf{\xi} = \begin{bmatrix} \mathbf{\theta}^T, \mathbf{\mu}^T \end{bmatrix}^T \mathbf{\xi}_0, \quad \mathbf{\theta}_0 \]

and \( \mathbf{\mu}_0 \) the true values of the parameters;

\[ \mathbf{Y} = [y(1) \cdots y(N)] ; \]

\[ \mathbf{S}(\mathbf{\theta}) = [s(1, \mathbf{\theta}) \cdots s(N, \mathbf{\theta})] \text{ the signal matrix ;} \]

\[ \mathbf{N} = [n(1) \cdots n(N)] ; \]

\[ \mathbf{Q}_0 = \mathbf{Q}(\mathbf{\mu}_0) \text{ and } \mathbf{S}_0 = \mathbf{S}(\mathbf{\theta}_0) . \]

Unless necessary, the dependence of \( \mathbf{S} \) and \( \mathbf{Q} \) on \( \mathbf{\theta} \) and \( \mathbf{\mu} \) 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:

\[ L(\mathbf{\xi}) = \ln |\mathbf{Q}| + N^{-1} \text{tr} \left( \mathbf{Q}^{-1} (\mathbf{Y} - \mathbf{S}) (\mathbf{Y} - \mathbf{S})^H \right) . \]

(3)

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

\[ \mathbf{S}(\mathbf{\theta}) = \mathbf{H}(\mathbf{\theta}) \mathbf{X} \quad (3.a) \]

where:

\[ \mathbf{H}(\mathbf{\theta}) \text{ an unknown } M \times q \text{ matrix ;} \]

\[ \mathbf{X} \text{ is a known } L \times q \text{ matrix.} \]

We also set:

\[ \mathbf{H}(\mathbf{\theta}_0) = \mathbf{H}_0 , \quad (3.b) \]

\[ \mathbf{H}(\mathbf{\theta}) = \mathbf{H} . \quad (3.c) \]

**Consistency:**

Hypothesizes:

\[ \text{H1: } \left\| \mathbf{Q}(\mathbf{\mu})^{-1} \right\|_F \text{ is bounded for all values of } \mathbf{\mu}. \]

\[ \text{H2: } \left\| \mathbf{H} \right\|_F \text{ is bounded for all values of } \mathbf{\theta}. \]
Lemma 1:

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

$$f(\xi) = \ln|Q| + tr(Q^{-1}Q_0) + tr(Q^{-1}(H-H_0)R_{xx}(H-H_0)^H).$$

(4)

Proof: The ML criterion (3) can be rewritten

$$L(\xi) = \ln|Q| + N^{-1}tr\left(Q^{-1}(Y-S)(Y-S)^H\right) - \ln|Q| + N^{-1}tr\left(Q^{-1}NN^H\right) + \Re\left(Q^{-1}(H-H_0)(N^{-1}XX^H)(H-H_0)^H\right) - 2\Re\left(tr(Q^{-1}(H-H_0)(N^{-1}XX^H))\right).$$

Assume,

$$\varepsilon(\xi) = L(\xi) - f(\xi) = \varepsilon_1(\xi) + \varepsilon_2(\xi) + \varepsilon_3(\xi)$$

(5)

where:

$$\varepsilon_1(\xi) = tr\left(Q^{-1}(N^{-1}NN^H - Q_0)\right),$$

$$\varepsilon_2(\xi) = tr\left(Q^{-1}(H-H_0)(N^{-1}XX^H - R_{xx})(H-H_0)^H\right),$$

$$\varepsilon_3(\xi) = -2\Re\left(tr(Q^{-1}(H-H_0)(N^{-1}XX^H))\right).$$

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

$$|\varepsilon(\xi)| \leq \left\|Q^{-1}\right\| \left\|N^{-1}NN^H - Q_0\right\|.$$  

From the weak law of large numbers, $\left\|N^{-1}BB^T - Q_0\right\|$ converges w.p.1 to 0. Since $\left\|Q^{-1}\right\|_F$ is bounded by H1, it follows that $\varepsilon_1(\xi)$ converges w.p.1 to 0 uniformly in $\xi$. Similarly,

$$|\varepsilon_2(\xi)| \leq \left\|(H-H_0)^HQ^{-1}(H-H_0)\right\|_F \left\|N^{-1}XX^H - R_{xx}\right\|_F$$

bounded from H1 and H2

The right hand side term of this inequality converges uniformly to 0 from H1, H2 and H3. Finally, it is easy to show that $\varepsilon_2(\xi)$ is zero-mean Gaussian with variance

$$\sigma^2 = 2N^{-1}tr\left[Q^{-1}Q_0Q^{-1}(H-H_0)(N^{-1}XX^H)(H-H_0)^H\right].$$

(6)

From H1 to H3, the term $tr\left[Q^{-1}Q_0Q^{-1}(H-H_0)(N^{-1}XX^H)(H-H_0)^H\right]$ in (6) is bounded, and thus $\sigma^2$ converges to 0 uniformly in $\xi$. Therefore $\varepsilon_2(\xi)$ converges w.p.1 to 0 uniformly in $\xi$. 

4216
Thus, $\mathcal{g}(\xi)$ converges w.p.1 to 0 uniformly in $\xi$. •

P.S: If we have

\[
\begin{align*}
 f_n(x, \alpha) & \xrightarrow{\text{uniformly inc}} f(\alpha) \\
 \arg\min_{\alpha} f(x) & = \theta
\end{align*}
\]

then $\theta_n = \arg\min_{\alpha} f_n(x, \alpha) \xrightarrow{\text{w.p.1}} \theta$ [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

\[
f(\xi) = \ln |Q| + tr(Q^{-1}Q_0) + tr\left(Q^{-1}(H - H_0) R_{\alpha}(H - H_0)^H\right)
\]

has a global minimum for $\theta = \theta_0$ and $\mu = \mu_0$. It is well known that the term $\ln |Q| + tr(Q^{-1}Q_0)$ in (7) is minimal if $Q = Q_0$, that is when $\mu = \mu_0$.

Moreover $tr\left(Q^{-1}(H - H_0) R_{\alpha}(H - H_0)^H\right)$ in (7) is positive, and from H4 it admits its minimal value 0 iff $\theta = \theta_0$.

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

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

2.3.1 Cramer Rao Bound (CRB):
The Fisher Information Matrix (FIM) for estimating $\xi = [\Theta^T, \mu^T]^T$ from $N$ snapshots is known and easily shown to be given by (cf. appendix A):

\[
F_\xi = \begin{bmatrix} F_\Theta & 0 \\ 0 & F_\mu \end{bmatrix}
\]

where:

\[
F_{\Theta}(i, j) = 2 \Re tr \left( \frac{\partial H}{\partial \Theta_i} \right)^H Q_0^{-1} \left( \frac{\partial H}{\partial \Theta_j} \right) X X^H \right), \quad \text{for} \quad 1 \leq i, j \leq P_{\Theta};
\]

\[
F_{\mu}(i, j) = N tr \left( Q_0^{-2} \frac{\partial Q}{\partial \mu_i} Q_0^{-1} \frac{\partial Q}{\partial \mu_j} \right), \quad \text{for} \quad 1 \leq i, j \leq P_{\mu};
\]

2.3.2 Efficiency:

Let us make the following hypothesis:

H5: $H$ and $Q$ are twice derivable with continuous derivatives at $\Theta_0$ and $\mu_0$

Theorem 2: Under hypothesizes H1 to H5 the MLE $\hat{\xi}$ is asymptotically Gaussian and efficient in the following sense:
\[ \sqrt{N} \left( \xi - \xi_0 \right) \] is asymptotically Gaussian, zero mean with covariance matrix \( V^{-1} \) where

\[ V = \lim_{N \to \infty} \frac{1}{N} F_N. \]  

\[ \textit{Proof}: \] First note that \( \lim_{N \to \infty} \frac{1}{N} F_N \) which is involved in the above theorem is well defined thanks to H3. Indeed, we have from (9):

\[ \frac{1}{N} F_N (i, j) = \frac{2}{N} \text{tr} \left[ \left( \begin{array}{cc} \partial H \partial \theta_j \end{array} \right) \left( \begin{array}{cc} \partial H \partial \theta_j \end{array} \right) \right] \xrightarrow{N \to \infty} 2 \text{tr} \left[ \left( \begin{array}{cc} \partial H \partial \theta_j \end{array} \right) \left( \begin{array}{cc} \partial H \partial \theta_j \end{array} \right) \right] R_{ij}. \]

and obviously:

\[ \frac{1}{N} F_N (i, j) = \text{tr} \left[ Q_0^{-1} \left( \begin{array}{cc} \partial \varphi \partial \mu_i \end{array} \right) \left( \begin{array}{cc} \partial \varphi \partial \mu_i \end{array} \right) \right]. \]

Thus, \( \lim_{N \to \infty} \frac{1}{N} F_N \) is well defined.

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

\[ \left( \frac{\partial L}{\partial \xi} \right) \approx \left( \frac{\partial L}{\partial \xi_0} \right) + \sum_i \left( \frac{\partial L}{\partial \xi \partial \xi_i} \right) \left( \xi_i - \xi_0 \right) = 0. \]

where

\[ \xi' = \gamma \xi_0 + (1 - \gamma) \hat{\xi} \] avec \( \gamma \in [0, 1] \)

\[ \Delta = \left( \begin{array}{c} \theta' \\ \mu' \end{array} \right) \]

Thus, \( \hat{\xi} - \xi_0 = -H(\xi')^{-1}G \).

where \( G = (G_i) \) and \( H(\xi') = H_p(\xi') \) denote respectively the gradient at \( \xi_0 \) and the Hessian at \( \xi' \)

\[ G_i = \left( \begin{array}{c} \partial L \\ \partial \xi_i \end{array} \right) \xi_0, \]

\[ H_{ij}(\xi') = \left( \begin{array}{c} \partial^2 L \\ \partial \xi_i \partial \xi_j \end{array} \right) \xi'. \]

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

\textit{Lemma 2:}

\( \sqrt{N} G \) where \( G \) is defined by (16), is asymptotically Gaussian with zero-mean and covariance matrix
Proof:
The components of the gradient with respect to $\mu_k$ are given by:

\[
\left( \frac{\partial L}{\partial \mu_k} \right) = tr \left( \frac{\partial Q}{\partial \mu_k} \right) Q_0^{-1} - N^{-1} tr \left( Q_0^{-1} \left( \frac{\partial Q}{\partial \mu_k} \right)_0 \right) = -N^{-1} tr \left( Q_0^{-1} \left( \frac{\partial Q}{\partial \mu_k} \right)_0 \right) \tag{18}
\]

Since $E[N BB^H] = Q_0$, we have: $E \left( \frac{\partial L}{\partial \mu_k} \right) = 0$

Similarly

\[
\left( \frac{\partial L}{\partial \theta} \right) = -N^{-1} tr \left( Q_0^{-1} \left( \frac{\partial H}{\partial \theta} \right)_0 \right) \tag{19}
\]

which implies, since $B$ is zero-mean: $E \left( \frac{\partial L}{\partial \theta} \right) = 0$

Thus, gradient $G$ is zero-mean. Furthermore, we see from expression (19) that $\left( \frac{\partial L}{\partial \theta} \right)$ is Gaussian, while the central limit theorem applied to $N^{-1} NN^H - N^{-1} \sum_{i=1}^N \mathbf{n}(t)\mathbf{n}(t)^H$ in (18) implies that $\sqrt{N} E \left( \frac{\partial L}{\partial \mu_k} \right)$ is asymptotically Gaussian: therefore $\sqrt{N} G$ is asymptotically Gaussian. Finally, a simple computation shows that the covariance of the elements $\sqrt{N} G$ is given by $\frac{1}{N} F_{G}$ which completes the proof of lemma 2 (cf. appendix B).

Lemma 3:

$N^{-1} XN^H \xrightarrow{p} 0$ as $N$ tends to infinity.

Proof:

The elements of $N^{-1} XN^H$ 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 $E \left( \left\| N^{-1} XN^H \right\|_F^2 \right)$ tends to zero as $N \rightarrow \infty$ which implies lemma 3. Indeed, it is easy to show that:

\[
E \left( \left\| N^{-1} XN^H \right\|_F^2 \right) = E \left[ tr \left( N^{-2} XN^H XN^H \right) \right] \leq N^{-1} tr \left( Q_0 \left( N^{-1} X X^H \right) \right)
\]
which tends to zero as \( N \to \infty \) from H3.

P.S: \( \|AB\|_F \leq \|A\|_F \|B\|_F \) and \( tr(AB) = tr(A)tr(B) \)

**Corollary:**

If \( \theta \overset{p}{\to} \theta_0 \) as \( N \to \infty \), then \( N^{-1} (Y - S)(Y - S)^H \overset{p}{\to} Q_0 \)

**Proof:** Let us split \( N^{-1} (Y - S)(Y - S)^H \) to:

\[
N^{-1} (Y - S)(Y - S)^H = N^{-1} (S_0 + N - S)(S_0 + N - S)^H = N^{-1} N N^H + (H - H_0)(N^{-1} X X^H)(H - H_0)^H - (H - H_0)(N^{-1} X N^H) - (N^{-1} N X^H)(H - H_0)^H
\]

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

\[
N^{-1} N N^H \overset{p}{\to} Q_0 \quad (21)
\]

In the second term \( (H - H_0)(N^{-1} X X^H)(H - H_0)^H \), the continuity of \( H \) (which results from H5) implies that

\[
H \overset{p}{\to} H_0 \quad \text{as} \quad \theta \overset{p}{\to} \theta_0 \quad (22)
\]

It follows from H3 that:

\[
(H - H_0)(N^{-1} X X^H)(H - H_0)^H \overset{p}{\to} 0 \quad \text{as} \quad N \to \infty.
\]

Finally, from lemma 3 and equation (22), the last two terms in (20) converge w.p.1 to 0.

**Lemma 4:**

The Hessian matrix \( \mathbf{H}(\xi) \) defined by (17) converges w.p.1 to \( \mathbf{V} \) (10.a).

**Proof:**

Let us study the behavior of the three terms \( \left( \frac{\partial^2 L}{\partial \mu \partial \mu} \right)_i \), \( \left( \frac{\partial^2 L}{\partial \mu \partial \theta} \right)_i \), and \( \left( \frac{\partial^2 L}{\partial \theta \partial \theta} \right)_i \) as \( N \to \infty \).

**Analysis of \( \left( \frac{\partial^2 L}{\partial \mu \partial \mu} \right)_i \):**

We have got:

\[
\left( \frac{\partial L}{\partial \mu_i} \right)_i = tr \left( Q^{-1} \frac{\partial Q}{\partial \mu_i} \right) - N^{-1} tr \left( Q^{-1} \frac{\partial Q}{\partial \mu_i} Q^{-1} (Y - S)(Y - S)^H \right).
\]

From which we deduce:
\[
\left( \frac{\partial^2 L}{\partial \mu_i \partial \mu_j} \right)_{\xi} = -\text{tr} \left( Q^{-1} \frac{\partial Q}{\partial \mu_i} Q^{-1} \frac{\partial Q}{\partial \mu_j} \right) + \text{tr} \left( Q^{-1} \frac{\partial^2 Q}{\partial \mu_i \partial \mu_j} \right) \\
-2N^{-1} \text{tr} \left( Q^{-1} \frac{\partial Q}{\partial \mu_i} Q^{-1} (Y - S)(Y - S)^H \right) \\
+2N^{-1} \text{tr} \left( Q^{-1} \frac{\partial Q}{\partial \mu_i} Q^{-1} \frac{\partial Q}{\partial \mu_j} Q^{-1} (Y - S)(Y - S)^H \right) \\
+2N^{-1} \text{tr} \left( Q^{-1} \frac{\partial Q}{\partial \mu_i} Q^{-1} \frac{\partial Q}{\partial \mu_j} Q^{-1} (Y - S)(Y - S)^H \right) \\
\] (24)

The consistence of MLE implies \( \hat{\xi} \xrightarrow{p} \xi_0 \).

where:
\[
\hat{\xi} = \gamma \hat{\xi}_0 + (1 - \gamma) \hat{\xi} \text{ with } \gamma \in [0, 1]
\]

Therefore, in the second member of (24) we have:
\[
Q \xrightarrow{p} Q_0 \text{and} \left( \frac{\partial^2 L}{\partial \mu_i \partial \mu_j} \right)_{\xi} \xrightarrow{p} \left( \frac{\partial^2 L}{\partial \mu_i \partial \mu_j} \right)_{\xi_0}
\]

Furthermore, according to the corollary of lemma 3, we have in the second member of (24):
\[
N^{-1} (Y - S)(Y - S)^H \xrightarrow{p} Q_0.
\]

Then:
\[
\left( \frac{\partial^2 L}{\partial \mu_i \partial \mu_j} \right)_{\xi} \xrightarrow{p} \text{tr} \left( Q_0^{-1} \left( \frac{\partial Q}{\partial \mu_i} Q_0^{-1} \left( \frac{\partial Q}{\partial \mu_j} \right) \right)_{\xi_0} \right)
\]

(25)

Analysis of \( \left( \frac{\partial^2 L}{\partial \mu_i \partial \theta_j} \right)_{\xi} \):
\[
\left( \frac{\partial^2 L}{\partial \mu_i \partial \theta_j} \right)_{\xi} = -2N^{-1} \text{tr} \left( Q^{-1} \frac{\partial Q}{\partial \mu_i} Q^{-1} \frac{\partial S}{\partial \theta_j} (S_0 - S + N)^H \right) + 2N^{-1} \text{tr} \left( Q^{-1} \frac{\partial Q}{\partial \mu_i} Q^{-1} (S_0 - S + N) \frac{\partial S}{\partial \theta_j} \right) \\
= 2N \left\{ \text{tr} \left( Q^{-1} \frac{\partial Q}{\partial \mu_i} Q^{-1} \frac{\partial H X X^H}{N} (H_0 - H)^H \right) \right\} + 2N \left\{ \text{tr} \left( Q^{-1} \frac{\partial Q}{\partial \mu_i} Q^{-1} \frac{\partial H}{\partial \theta_j} X X^H N^{-1} \right) \right\}
\] (26)

We easily verify that the two terms of the second member in (26) converge in probability to 0 by utilizing hypothesis H3, lemma 3 and \( \hat{\xi} \xrightarrow{p} \xi_0 \):
\[
\left( \frac{\partial^2 L}{\partial \mu_i \partial \theta_j} \right)_{\xi} \xrightarrow{p} 0
\]

(27)
Analysis of $\frac{\partial^2 L}{\partial \theta \partial \theta_j}$

\[
\left( \frac{\partial^2 L}{\partial \theta_i \partial \theta_j} \right)_{\xi} = 2N^{-1} \text{tr} \left( \frac{\partial^2 \mathbf{S}^H}{\partial \theta_i \partial \theta_j} Q^{-1} \frac{\partial \mathbf{S}}{\partial \theta_j} - Q^{-1} \frac{\partial^2 \mathbf{S}}{\partial \theta_i \partial \theta_j} (s_0 - s + \mathbf{N})^H \right)
\]

\[
= 2N^{-1} \text{tr} \left( X^H \left( \frac{\partial \mathbf{H}}{\partial \theta_j} \right)^H Q^{-1} \left( \frac{\partial \mathbf{H}}{\partial \theta_j} \right) X - Q^{-1} \frac{\partial^2 \mathbf{H}}{\partial \theta_i \partial \theta_j} X (H_0 - \mathbf{H}) X + \mathbf{N}^H \right)
\]

\[
= 2 \text{ tr} \left( \frac{XX^H}{N} \left( \frac{\partial \mathbf{H}}{\partial \theta_i} \right)^H Q^{-1} \left( \frac{\partial \mathbf{H}}{\partial \theta_i} \right) - 2 \text{ tr} \left( Q^{-1} \frac{\partial^2 \mathbf{H}}{\partial \theta_i \partial \theta_j} \left( \frac{XX^H}{N} (H_0 - \mathbf{H})^H + \frac{\mathbf{N}^H}{N} \right) \right) \right)
\]

(28)

We easily verify that the two terms of the second member in (28) converge in probability to 0 by utilizing hypothesis H3, lemma 3 and £ $\rightarrow \xi_0$.

\[
\left( \frac{\partial L}{\partial \theta_i \partial \theta_j} \right)_{\xi} \rightarrow 0
\]

(29)

Thus, $\mathbf{H}(0)$ converges w.p.1 to $\mathbf{V} = \lim_{N \rightarrow \infty} \frac{1}{N} \mathbf{F}_{\xi}$.

Finally, lemmas 2, 4, and the expression (15) demonstrate theorem 2.

**Cramer-Rao bounds (CRB):**

We calculate in this paragraph the CRB of the two models 1 and 2 (given below) of the impulse response $\mathbf{H}(0)$ presented in (3.a) using the general expression (9). Note that the obtained expression of CRB for the model 2 is original.

From the general expression of received signals (1):

\[
\mathbf{s}(t, \theta) = \mathbf{H}(\theta) \mathbf{x}(t)
\]

(30)

Where $\theta$ designates all the parameters of $\mathbf{H}$. unless necessary, the dependence of $\mathbf{H}$ on $\theta$ will be omitted.

Let us recall briefly some matrices operations:

A new operation « the Kronecker product » noted by $\otimes$ and defined by the following way:

Let $\mathbf{A}$ and $\mathbf{B}$ be two matrices composed of $M$ lines and whose $m^{th}$ lines are $\mathbf{a}_m$ and $\mathbf{b}_m$ respectively:

\[
\mathbf{A} = \begin{bmatrix} a_1 \\ \vdots \\ a_M \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} b_1 \\ \vdots \\ b_M \end{bmatrix}
\]

Then $\mathbf{A} \otimes \mathbf{B}$ designates the following matrix:

\[
\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} a_1 \otimes b_1 \\ \vdots \\ a_M \otimes b_M \end{bmatrix}
\]

Let us start by the less refined model.
3.1 Model 1:

The impulse response $H$ is globally considered unknown. The FIM for this problem is well known. Let us recall briefly the principal results.

The unknown parameters vector can be written as:

$$\theta = \begin{bmatrix} \text{vec} \left[ \Re(H) \right] \\ \text{vec} \left[ \Im(H) \right] \end{bmatrix}$$

(31)

The corresponding FIM is:

$$F_{\theta} = \begin{bmatrix} F_{11} & F_{12}^T \\ F_{21} & F_{22} \end{bmatrix}$$

(32)

where:

$$F_{11} = 2 \Re \left( \left[ X \right]^T \otimes Q^{-1} \right).$$

(33)

and

$$F_{12} = 2 \Im \left( \left[ X \right]^T \otimes Q^{-1} \right).$$

(34)

P.S.: let $z = x + jy$ ($x = \Re(z)$ and $y = \Im(z)$) be a zero-mean circular random complex vector:

$$E[z z^+] = 0$$

Then:

$$\text{cov} \left[ \begin{bmatrix} x \\ y \end{bmatrix} \right] = \begin{bmatrix} R_{xx} & -R_{yx} \\ R_{yx} & R_{yy} \end{bmatrix} = \begin{bmatrix} R_{xx} & -R_{yx} \\ R_{yx} & R_{yy} \end{bmatrix}.$$  

(35)

where $R_{xx}$ is asymmetric.

We verify that the inverse of the covariance matrix has a similar structure:

$$\begin{bmatrix} R_{xx} & -R_{yx} \\ R_{yx} & R_{yy} \end{bmatrix}^{-1} = \begin{bmatrix} C & -D \\ D & C \end{bmatrix}.$$  

(36)

The covariance of $z$ is then expressed by:

$$E[(x + jy)(x + jy)^H] = 2 (R_{xx} + jR_{yx}).$$  

(37)

And we show that:

$$R_{xx} + jR_{yx} = (C + jD)^{-1}.$$  

(38)

We verify that we have in expression (32) of the FIM: $F_{21}^T = -F_{21}$.

Therefore, the FIM has an expression of the form (36). It follows that if we have an efficient estimator $\hat{\theta}$ then the complex vector $\hat{\theta} = \text{vec} \left[ \Re(H) \right] + j \text{vec} \left[ \Im(H) \right]$ is circular of covariance:

$$E \left[ (\hat{\theta} - \theta)(\hat{\theta} - \theta)^H \right] = 2 (F_{11} + jF_{12})^{-1} - \left( \left[ X \right]^T \otimes Q \right)^{-1}.$$  

(39)
The expression (39) permits us to work with a vector of complex parameters of reduced dimension of factor of two.

**Model 2:**

Let us consider now that the $Mxq$ matrix $H$ in (30) is an unknown complex matrix of a rank $J$ so that:

$$\mathbf{H} = \mathbf{AB}$$  \hspace{1cm} (40)

where:

$A$ ($M'J$) is an arbitrary complex matrix;

$\mathbf{B} = [\mathbf{I} \quad \mathbf{B}_0]$ ($J'q$) where $\mathbf{B}_0$ ($J'(q-J)$) is an arbitrary matrix.

The vector $\theta$ of unknown parameters in $H$ is written as:

$$\theta = \begin{bmatrix}
\text{vec} \left[ \Re(\mathbf{A}) \right] \\
\text{vec} \left[ \Im(\mathbf{A}) \right] \\
\text{vec} \left[ \Re(\mathbf{B}_0) \right] \\
\text{vec} \left[ \Im(\mathbf{B}_0) \right]
\end{bmatrix}$$  \hspace{1cm} (41)

We show in appendix C that the FIM has the following expression:

$$\mathbf{F}_\theta = \begin{bmatrix}
\mathbf{F}_{11} & \mathbf{F}_{12}
\\
\mathbf{F}_{21} & \mathbf{F}_{11}
\end{bmatrix}$$  \hspace{1cm} (42)

where:

$$\mathbf{F}_{11} = \begin{bmatrix}
\Re(\mathbf{A}) & \Re(\mathbf{C}) \\
\Re(\mathbf{C}^T) & \Re(\mathbf{B})
\end{bmatrix}$$  \hspace{1cm} (43)

and

$$\mathbf{F}_{21} = \begin{bmatrix}
\Im(\mathbf{A}) & \Im(\mathbf{C}) \\
-\Im(\mathbf{C}^T) & \Im(\mathbf{B})
\end{bmatrix}.$$  \hspace{1cm} (44)

with:

$$\mathbf{A} = 2 \left[ \mathbf{B} \mathbf{XX}^H \mathbf{B}^H \right]^T \otimes \mathbf{Q}^{-1}.$$  \hspace{1cm} (45)

$$\mathbf{B} = 2 \left[ \begin{bmatrix}
\mathbf{0} \\
\mathbf{I}_{q-J}
\end{bmatrix} \mathbf{XX}^H \mathbf{I}_{q-J} \right]^T \otimes \left[ \mathbf{A}^H \mathbf{Q}^{-1} \mathbf{A} \right].$$  \hspace{1cm} (46)

$$\mathbf{C} = 2 \left[ \begin{bmatrix}
\mathbf{0} \\
\mathbf{I}_{q-J}
\end{bmatrix} \mathbf{XX}^H \mathbf{B}^H \mathbf{B}^H \mathbf{B}^H \mathbf{X} \mathbf{X}^H \mathbf{B} \mathbf{B}^H \right]^T \otimes \left[ \mathbf{Q}^{-1} \mathbf{A} \right].$$  \hspace{1cm} (47)

We verify that we have in the expression (42) of the FIM:

$$\mathbf{F}_{21} = -\mathbf{F}_{21}. \hspace{1cm} (48)$$

Therefore, the FIM has an expression of the form (36). It follows that we have the use of an efficient estimator

$$\hat{\theta} = -\frac{\text{vec} \left[ \Re(\hat{\mathbf{A}}) \right] + \text{vec} \left[ \Im(\hat{\mathbf{A}}) \right]}{\text{vec} \left[ \Re(\hat{\mathbf{B}_0}) \right] + \text{vec} \left[ \Im(\hat{\mathbf{B}_0}) \right]}$$

is circular of covariance.
The expression (49) permits us to work with a vector of complex parameters of reduced dimension of factor of two.

P.S: The MLE of the multi-channel impulse response $H$ in a mobile communication system whose base stations equipped with antennas arrays with a reduced rank (model 2) has been done in our previous paper (Kassem and Fortser, 2003).

Applications to calculate the MSE of $H$:

It is interesting to have in practice a global indicator of the estimation precision of the channel $H$. A such indicator is provided by the Mean Square Error.

$$MSE = E \| \hat{H} - H \|^2_f$$

Suppose that we have the use of an estimator asymptotically consistent and efficient (as for example the ML estimator of the channel $H$), we demonstrate in this paragraph the following results.

**Theorem 3:**

When the number of observations $N$ tends to infinity,

$$E \| \hat{H}(\hat{\theta}) - H \|^2_f = \sum_i \sum_j \Re \text{tr} \left[ \frac{\partial H}{\partial \theta_i} \frac{\partial H^H}{\partial \theta_j} \right] V_{ij},$$

where $V = F^{-1}$ is the inverse of the FIM.

**Proof:**

When $N$ tends to infinity, the consistency of $\hat{\theta}$ shows it is close to $\theta_0$. A first order Taylor expansion gives then:

$$\hat{H}(\hat{\theta}) = H(\theta) + \Delta H = H(\theta) + \sum_{p=1}^{P} \left( \frac{\partial H}{\partial \theta_p} \right) (\hat{\theta}_p - \theta_0, \theta_p).$$

and:

$$\| \hat{H}(\hat{\theta}) - H \|^2_f = \text{tr} \left[ \Delta H \Delta H^H \right]$$

$$= \sum_{i=1}^{P} \sum_{j=1}^{P} \text{tr} \left[ \frac{\partial H}{\partial \theta_i} \frac{\partial H^H}{\partial \theta_j} \right] (\hat{\theta}_i - \theta_0, \theta_i)(\hat{\theta}_j - \theta_0, \theta_j)$$

$\hat{\theta}$ is being asymptotically efficient, we have:

$$E \left[ (\hat{\theta}_i - \theta_0, \theta_i)(\hat{\theta}_j - \theta_0, \theta_j) \right] = V_{ij},$$

which conducts to:

$$E \| \Delta H \|^2_f = \sum_{i=1}^{P} \sum_{j=1}^{P} \text{tr} \left[ \frac{\partial H}{\partial \theta_i} \frac{\partial H^H}{\partial \theta_j} \right] V_{ij},$$

which can be written again, since $E \| \Delta H \|^2_f$ is real:
Theorem 4:
For the models of used signal of the form (30), the MSE can be written as:
\[ E\|\hat{H} - H\|_F^2 = \frac{1}{2} \text{tr}[G V], \]  
where:
\[ G = \text{obtained by replacing } Q \text{ and } XX'' \text{ by the identity matrix } I \text{ in the FIM (9)}. \]
\[ V = F^{-1}_\theta \]

Proof:
It is a direct consequence of the expression (50) and of the form (9) of the FIM.

Theorem 5:
Suppose that the FIM has the following form (cf. models 1 and 2):
\[ F_\theta = \begin{bmatrix} F_{11} & -F_{31} \\ F_{21} & F_{11} \end{bmatrix}, \]  
where the matrices \( F_{ij} \) are squared of the same dimension. Then the matrix \( G \) defined in theorem 4 has a similar structure:
\[ G = \begin{bmatrix} G_{11} & -G_{21} \\ G_{21} & G_{11} \end{bmatrix} \]
and the MSE (52) can be written as:
\[ E\|\hat{H} - H\|_F^2 = \text{tr} \left( G_{11} + jG_{21} \right) \left( F_{11} + jF_{21} \right)^{-1}. \]

As we will see in the following paragraph, theorem 5 permits to calculate easily the MSE \( E\|\hat{H} - H\|_F^2 \) for the models 1 and 2.

Proof:
According to theorem 4, we have:
\[ E\|\hat{H} - H\|_F^2 = \frac{1}{2} \text{tr}[G V]. \]
Since \( G \) is obtained from \( F \) by replacing \( Q \) and \( XX'' \) by the identity \( I \), and that \( V = F^{-1}_\theta \), the structure (53) of \( F \) is preserved in \( G \) and \( V \).
\[ G = \begin{bmatrix} G_{11} & -G_{21} \\ G_{21} & G_{11} \end{bmatrix} \]
and
\[ V = \begin{bmatrix} V_{11} & -V_{21} \\ V_{21} & V_{11} \end{bmatrix} \]
Thus, the expression (52) can be written as:

$$E\|\hat{H} - H\|^2 = \text{tr} [G_{v1}V_{v1} - G_{v2}V_{v2}]$$

which can be written again:

$$E\|\hat{H} - H\|^2 = \text{tr} \left\{ G_{v1} + jG_{v2} \right\} \left[ V_{v1} + jV_{v2} \right]$$

Again, taking into account the inequality (38)

$$E\|\hat{H} - H\|^2 = \text{tr} \left\{ G_{v1} + jG_{v2} \right\} \left[ V_{v1} + jV_{v2} \right]$$

(58)

**Calculation of $E\|\hat{H} - H\|^2$ for the model 1:**

The expression (32) shows that we are in the hypotheses of theorem 5. From the expressions (33) and (34) we have:

$$G_{11} = 2I$$
$$G_{21} = 0,$$

$$F_{11} + jF_{21} = 2(XX^H)^T \otimes Q^{-1}.$$  

We deduce:

$$E\|\hat{H} - H\|^2 = \text{tr} \left[ (XX^H)^T \right] \text{tr} [Q].$$

(59)

**Calculation of $E\|\hat{H} - H\|^2$ for the model 2:**

The expression (42) shows that we are in the hypotheses of theorem 5. From the expressions (43) and (44) we have:

$$G_{11} + jG_{21} = 2 \begin{bmatrix} (BB^H)^T \otimes I_{J} & \left[ 0 \ I_{J-1} \right] B^H \otimes A \\ \left[ 0 \ I_{J-1} \right] B^H \otimes A^H & I_{J-1} \otimes A^H A \\ \end{bmatrix}$$

(60)

Assume in (58) that:

$$2(F_{11} + jF_{21})^{-1} = \begin{bmatrix} V_{aa} & V_{a\phi} \\ V_{a\phi} & V_{\phi \phi} \end{bmatrix},$$

(61)

which is the covariance (49) of the complex parameter $\hat{\theta}_e$

The matrices $V_{aa}$, $V_{a\phi}$ and $V_{\phi \phi}$ are given by (by utilizing the inverse by blocs to calculate the expression 49):

$$V_{aa} = (A - C B^{-1} C^H)^{-1},$$

$$V_{a\phi} = (B - C^H A^{-1} C)^{-1},$$

$$V_{\phi \phi} = - \left( A - C B^{-1} C^H \right)^{-1} C B^{-1},$$

$$= - V_{aa} C B^{-1}$$

(62) (63) (64)
where the matrices \( \mathbf{A}, \mathbf{B} \) and \( \mathbf{C} \) are defined by (45), (46) and (47).

The expression (58) of the MSE conducts to (according to (60) and (61)):

\[
\mathbb{E} \left\| \hat{\mathbf{H}} - \mathbf{H} \right\|^2 = \text{tr} \left[ (\mathbf{I}_{q-j} \otimes \mathbf{A}^H \mathbf{A}) \mathbf{V}_{k_h} \right] + \text{tr} \left[ (\mathbf{B}^T \mathbf{B}^T \otimes \mathbf{I}_{k_e}) \mathbf{V}_{e_b} \right] + 2 \Re \left\{ \text{tr} \left[ \left( [\mathbf{0} \quad \mathbf{I}_{q-j}] \mathbf{B}^T \right) \otimes \mathbf{A}^H \right) \mathbf{V}_{e_b} \right\}
\]

(65)

4. Simulations:
We present in this paragraph some results of simulation which show the importance of taking into account a multipath propagation model (refined model) of the channel \( \mathbf{H} \), thus the efficiency of the method ML for the models 1 and 2, the MLE has an analytic solution.

Our indicator is the MSE of the channel \( \mathbb{E} \left\| \hat{\mathbf{H}} - \mathbf{H} \right\|^2 \) which is given by the expressions (59) (for the model 1) and (65) (for the model 2).

The antenna array is uniform and linear (ULA), and it is equipped of \( M = 8 \) sensors omni-directional, the path directions are referred by their electric angles \( \alpha = 2 \pi \sin \alpha \frac{d}{\lambda} \) where \( d \) is the distance inter-sensors, is the wavelength, and \( \alpha \) is the path direction with respect to the antenna traverse.

The matrix \( \mathbf{H} \) of the impulse response has been generated according to:

\[
\mathbf{H} = \mathbf{A}(\alpha) \mathbf{C}
\]

where \( \mathbf{A}(\alpha) \) is the matrix of directional vectors of \( J \) paths \( \mathbf{\alpha}_1, \ldots, \mathbf{\alpha}_j \), with \( \mathbf{\alpha} = [\mathbf{\alpha}_1^T \cdots \mathbf{\alpha}_j^T]^T \) The elements of the matrix \( \mathbf{C} \) have been generated according to independent complex normal law, and the \( k \)th line of \( \mathbf{C} \) has been normalized in order his squared norm is equal to the power of the \( k \)th path.

In all the simulations (fig. 1, 2, 3 and 4), the covariance matrix \( \mathbf{Q} \) is equal to the identity, but, of course it is estimated in the simulations. The matrix of emitted signals \( \mathbf{X} \) is orthogonal, with:

\[
\frac{1}{N} \mathbf{X} \mathbf{X}^H = \mathbf{I}
\]

In all the cases, we note that, more \( N \) increases, more the MSE of the two models becomes close to the CRB. That is the efficiency of ML.

We also note that the MSE of the model 1 is always greater than the one of the model 2, which shows the interest of the refined model of the channel.

Finally, for a given number of paths, the CRB of models 1 and 2 are independent of the paths distance and power: it is clear for the model 1, because the FIM depends only on \( \mathbf{X} \) and \( \mathbf{Q} \) (32), and for the model 2, we show in appendix D that is the case when \( \mathbf{Q} \) and \( \mathbf{X} \mathbf{X}^H \) are proportional to identity \( \mathbf{I} \). However, the ML reaches with difficult the CRB when the powers of paths are low or when the paths are close to each other.

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 to the same law of probability, and we have calculated the corresponding CRB. The general results have been applied to the calculation of the MSE of different models of the channel \( \mathbf{H} \).

Furthermore, we have shown and verified using simulations: the importance of a refined model of the channel, the efficiency of the ML method, and the independence of MSE of the estimated channel with and without the rank constraint (models 2 and 1 respectively) of the direction and power of paths.
Fig. 1: MSE of the estimated channel (two paths): CRB and ML

Fig. 2: MSE of the estimated channel (two paths): CRB and ML

Fig. 3: MSE of the estimated channel (three paths): CRB and ML
Appendices:
Appendix A: Calculation of the FIM:

We demonstrate in this appendix the result given in the paragraph 2.3.1.

Unless necessary, the dependence of s on θ will be omitted, that is \( s(t) = s(t, \theta) \)

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

\[
f(y(t)) = \frac{1}{|Q|^N} \exp \left[ -\frac{1}{2} (y(t) - s(t))^H Q^{-1} (y(t) - s(t)) \right].
\]

which conducts for the all \( N \) independent observations to:

\[
f(y(1), \ldots, y(N)) = \prod_{t=1}^{N} f(y(t)) = \frac{1}{(|Q|)^N} \exp \left[ -\frac{1}{2} \sum_{t=1}^{N} (y(t) - s(t))^H Q^{-1} (y(t) - s(t)) \right].
\]

The corresponding log-likelihood is:

\[
L = \ln(f(y(1), \ldots, y(N))) = -N \ln|Q| - \sum_{t=1}^{N} (y(t) - s(t))^H Q^{-1} (y(t) - s(t)). \tag{A.1}
\]

Partial derivatives:

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

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

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

\[
= \sum_t y(t)^H Q^{-1} \frac{\partial s(t)}{\partial \theta_p} + \sum_t \frac{\partial}{\partial \theta_p} s(t)^H Q^{-1} y(t) - \sum_t \frac{\partial}{\partial \theta_p} s(t)^H Q^{-1} s(t) - \sum_t s(t)^H Q^{-1} \frac{\partial}{\partial \theta_p} s(t). \tag{A.2-a}
\]

\[
= 2\Re \left\{ \sum_t y(t)^H Q^{-1} \frac{\partial}{\partial \theta_p} s(t) - \sum_t s(t)^H Q^{-1} \frac{\partial}{\partial \theta_p} \bar{s}(t) \right\}. \tag{A.2-b}
\]
Utilizing the identities of matrices (Scharf, 1990):

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

and

$$\frac{\partial Q^{-1}}{\partial \mu_k} = -Q^{-1} \frac{\partial Q}{\partial \mu_k} Q^{-1}.$$

Therefore, the expression of partial derivatives of $L$ with respect to the parameters $\mu_k$ describing the covariance $Q$ of the noise:

$$\frac{\partial L}{\partial \mu_k} = -N \text{tr} \left( Q^{-1} \frac{\partial Q}{\partial \mu_k} \right) + \sum_t (y(t) - s(t))^T Q^{-1} \frac{\partial Q}{\partial \mu_k} Q^{-1} (y(t) - s(t)).$$

(A.3)

Calculation of the FIM:

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

$$F = \begin{bmatrix}
\frac{\partial L}{\partial \theta_1} & \frac{\partial L}{\partial \theta_2} & \cdots & \frac{\partial L}{\partial \theta_p} \\
\frac{\partial L}{\partial \theta_2} & \frac{\partial L}{\partial \theta_1} & \cdots & \frac{\partial L}{\partial \theta_p} \\
& \ddots & \ddots & \vdots \\
\frac{\partial L}{\partial \theta_p} & \frac{\partial L}{\partial \theta_p} & \cdots & \frac{\partial L}{\partial \theta_1}
\end{bmatrix}$$

(A.4)

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

and

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

Then we have to calculate the different terms of $F$.

Calculation of $E \left[ \frac{\partial^2 L}{\partial \theta \partial \theta} \right]$.
Starting from (A.2-a), we obtain:

\[
\frac{\partial^2 L}{\partial \theta_i \partial \theta_j} = \left( \sum_t y(t)^H Q^{-1} \frac{\partial^2 s(t)}{\partial \theta_i \partial \theta_j} \right) + \left( \sum_t y(t)^H Q^{-1} \frac{\partial^2 s(t)}{\partial \theta_i \partial \theta_j} \right)^H \\
+ \left( - \sum_t \frac{\partial^2 s(t)}{\partial \theta_i \theta_j} Q^{-1} s(t) - \sum_t \frac{\partial s(t)}{\partial \theta_j} Q^{-1} \frac{\partial s(t)}{\partial \theta_j} \right) \\
+ \left( - \sum_t \frac{\partial^2 s(t)}{\partial \theta_i \theta_j} Q^{-1} s(t) - \sum_t \frac{\partial s(t)}{\partial \theta_j} Q^{-1} \frac{\partial s(t)}{\partial \theta_j} \right)^H
\]  

(A.5)

Taking the expectation of the precedent expression, we obtain:

\[
E \left[ \frac{\partial^2 L}{\partial \theta_i \partial \theta_j} \right] = \sum_t s(t)^H Q^{-1} \frac{\partial^2 s(t)}{\partial \theta_i \partial \theta_j} + \sum_t \frac{\partial s(t)}{\partial \theta_i} Q^{-1} \frac{\partial s(t)}{\partial \theta_j} \\
- \sum_t \frac{\partial^2 s(t)}{\partial \theta_i \theta_j} Q^{-1} s(t) - \sum_t \frac{\partial s(t)}{\partial \theta_i} Q^{-1} \frac{\partial s(t)}{\partial \theta_j} \\
- \sum_t s(t)^H Q^{-1} \frac{\partial^2 s(t)}{\partial \theta_i \partial \theta_j} - \sum_t \frac{\partial s(t)}{\partial \theta_j} Q^{-1} \frac{\partial s(t)}{\partial \theta_i} \\
= -2 \Re \left[ \sum_t \frac{\partial s(t)}{\partial \theta_i} Q^{-1} \frac{\partial s(t)}{\partial \theta_j} \right] \\
= -2 \left[ \sum_t \frac{\partial s(t)}{\partial \theta_i} Q^{-1} \frac{\partial s(t)}{\partial \theta_j} \right]
\]  

(A.6)

Calculation of \( E \left[ \frac{\partial^3 L}{\partial \mu_k \partial \mu_k} \right] \)

Starting from (A.3), we have:

\[
\frac{\partial L}{\partial \mu_k} = - \frac{N}{m} \left( Q^{-1} \frac{\partial Q}{\partial \mu_k} \right) + \sum_t \left( y(t) - s(t) \right)^H Q^{-1} \frac{\partial Q}{\partial \mu_k} Q^{-1} \left( y(t) - s(t) \right) \\
= - \frac{N}{m} \left( Q^{-1} \frac{\partial Q}{\partial \mu_k} \right) + \sum_t \left( y(t) Q^{-1} \frac{\partial Q}{\partial \mu_k} Q^{-1} \right) \\
\]

(A.7)

thus,
Taking the expectation of the precedent expression, we obtain

\[
E \left[ \frac{\partial^2 L}{\partial \mu_i \partial \mu_j} \right] = -N \text{tr} \left( Q^{-1} \frac{\partial^2 Q}{\partial \mu_i \partial \mu_j} \right) + N \text{tr} \left( Q^{-1} \frac{\partial Q}{\partial \mu_i} Q^{-1} \frac{\partial Q}{\partial \mu_j} \right) + N \text{tr} \left( Q^{-1} \frac{\partial^2 Q}{\partial \mu_i \partial \mu_j} \right)
\]

\[
= -N \text{tr} \left( Q^{-1} \frac{\partial Q}{\partial \mu_i} Q^{-1} \frac{\partial Q}{\partial \mu_j} \right) - N \text{tr} \left( Q^{-1} \frac{\partial Q}{\partial \mu_i} \frac{\partial Q}{\partial \mu_j} \right)
\]

(A.9)

Calculation of \( E \left[ \frac{\partial^2 L}{\partial \theta_i \partial \mu_j} \right] \)

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

\[
\frac{\partial L}{\partial \theta_p} = 2\Re \left\{ \sum_t (y(t) - s(t))^H \frac{\partial s(t)}{\partial \theta_p} \right\},
\]

thus

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

(A.10)

which conducts to

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

(A.11)

**Final Result:**

The FIM can be written using the terms calculated previously:

\[
F = \begin{bmatrix}
F_x & 0 \\
0 & F_p
\end{bmatrix}
\]

(A.12)
where:
\[
F_g (i, j) = 2 \Re \sum_{t=1}^{N} \left( \frac{\partial s_j(t, \theta)}{\partial \theta_i} \right)^H Q^{-1} \left( \frac{\partial s_j(t, \theta)}{\partial \theta_j} \right), \quad \text{for } 1 \leq i, j \leq P_g .
\]
\[
F_\mu (i, j) = N \text{tr} \left( Q^{-1} \left( \frac{\partial Q}{\partial \mu_i} \right) Q^{-1} \left( \frac{\partial Q}{\partial \mu_j} \right) \right), \quad \text{for } 1 \leq i, j \leq P_\mu .
\]

Appendix B: Calculation of the covariance of \( \sqrt{N} G \)

We calculate in this appendix the covariance of \( \sqrt{N} G \) where the Gradient \( G \) is defined by (16):

\[
G_i = \left( \frac{\partial L}{\partial \xi_i} \right)_{\xi_i}
\]

Analysis of \( \frac{\partial L}{\partial \mu_j} \)

Starting from (18), we have:
\[
\left( \frac{\partial L}{\partial \mu_j} \right)_{\xi_i} = -\text{tr} \left( Q^{-1} \left( \frac{\partial Q}{\partial \mu_j} \right) Q^2 \left( \frac{1}{N} N N^t - Q_0 \right) \right).
\]

and
\[
N E \left[ \frac{\partial L}{\partial \mu_i} \frac{\partial L}{\partial \mu_j} \right] = N E \left[ \text{tr} \left( Q^{-1} \left( \frac{\partial Q}{\partial \mu_i} \right) Q^2 \left( \frac{1}{N} N N^t - Q_0 \right) \right) \text{tr} \left( Q^{-1} \left( \frac{\partial Q}{\partial \mu_j} \right) Q^2 \left( \frac{1}{N} N N^t - Q_0 \right) \right) \right]
\]

where \( N \) is circular complex Gaussian, then (Van der Vaart, 1998):
\[
E \left[ \text{tr} \left( \hat{Q} - Q_0 \right) A \text{tr} \left( \hat{Q} - Q_0 \right) B \right] = \frac{1}{N} \text{tr} \left[ Q_0 A Q_0 B \right].
\]

We deduce:
\[
N E \left[ \frac{\partial L}{\partial \mu_i} \frac{\partial L}{\partial \mu_j} \right] = \text{tr} \left( Q^{-1} \left( \frac{\partial Q}{\partial \mu_i} \right) Q^2 Q_0 Q^{-1} \left( \frac{\partial Q}{\partial \mu_j} \right) Q^2 \right)
\]
\[
= \text{tr} \left( Q^{-1} \left( \frac{\partial Q}{\partial \mu_j} \right) Q^2 \left( \frac{\partial Q}{\partial \mu_i} \right) \right)
\]

Analysis of \( \frac{\partial L}{\partial \mu_i} \)

\[
N E \left[ \frac{\partial L}{\partial \mu_i} \frac{\partial L}{\partial \mu_j} \right] = 0
\]

Moments of odd order of \( N \)

Analysis de \( \frac{\partial L}{\partial \xi_i} \)

4234
Starting from (19), we have:

\[
\left( \frac{\partial L}{\partial \theta'} \right)_\text{e} = -N^{-1} \left\{ \sum_{i=1}^{N} \mathbf{n}(t_i)^H Q_0^{-1} \left( \frac{\partial s(t_i)}{\partial \theta'} \right)_\text{e} + \sum_{i} \left( \frac{\partial s(t_i)^H}{\partial \theta'} \right)_\text{e} Q_0^{-1} \mathbf{n}(t_i) \right\}
\]

and:

\[
\text{N.E.} \left[ \frac{\partial L}{\partial \theta_i} \right] = N^{-1} \left\{ \sum_{i} \mathbb{E} \left[ \mathbf{n}(t_i)^H Q_0^{-1} \left( \frac{\partial s(t_i)}{\partial \theta_i} \right)_\text{e} \mathbf{n}(t_i) \right] \right\}
\]

\[
+ N^{-1} \left\{ \sum_{i} \mathbb{E} \left[ \mathbf{n}(t_i)^H Q_0^{-1} \left( \frac{\partial s(t_i)^H}{\partial \theta_i} \right)_\text{e} Q_0^{-1} \mathbf{n}(t_i) \right] \right\}
\]

\[
+ N^{-1} \left\{ \sum_{i} \mathbb{E} \left[ \left( \frac{\partial s(t_i)^H}{\partial \theta_i} \right)_\text{e} Q_0^{-1} \left( \frac{\partial s(t_i)^H}{\partial \theta_i} \right)_\text{e} \right] \right\}
\]

\[
= N^{-1} \left\{ \sum_{i} \left( \frac{\partial s(t_i)^H}{\partial \theta_i} \right)_\text{e} Q_0^{-1} \left( \frac{\partial s(t_i)^H}{\partial \theta_i} \right)_\text{e} \right\}
\]

\[
= 2N^{-1} \left\{ \sum_{i} \left( \frac{\partial s(t_i)^H}{\partial \theta_i} \right)_\text{e} Q_0^{-1} \left( \frac{\partial s(t_i)^H}{\partial \theta_i} \right)_\text{e} \right\}^{\frac{1}{2}}
\]

\[
= \frac{1}{N} \text{F}_\theta
\]

We deduce that the covariance of $\sqrt{N} \text{G}$ is given by the inverse of the FIM (8).

**Appendix C: Calculation of CRB for the model 2:**

To evaluate $\text{F}_\theta(9)$, we need to calculate the partial derivatives of the signal carrying the signal (30) with respect to the unknown parameters (41).

Assume: $\Re(A) = \{\alpha(i,j)\}$ $\Im(A) = \{\beta(i,j)\}$ $\Re(B) = \{\gamma(i,j)\}$ $\Im(B) = \{\delta(i,j)\}$

According to the expression (41), the unknown parameters are:

* $\alpha(i,j)$ and $\beta(i,j)$ for $1 \leq i \leq M$ and $1 \leq j \leq J$.
* $\gamma(i,j)$ and $\delta(i,j)$ for $1 \leq i \leq M$ and $J < j \leq Q$.

We have in the expression (A.13):

\[
\frac{\partial s(t)}{\partial \alpha(i,j)} = \mathbf{u}_i \mathbf{u}_j^H \mathbf{x}(t) \quad \text{and} \quad \frac{\partial s(t)}{\partial \beta(i,j)} = \mathbf{j}_i \mathbf{u}_j^H \mathbf{x}(t)
\]
\[
\begin{align*}
\frac{\partial s(t)}{\partial x(i,j)} & = x(t)^\dagger B^HBu_j^T \\
\frac{\partial s(t)}{\partial \gamma(i,j)} & = A_iu_j^TX(t) \\
\frac{\partial s(t)}{\partial \gamma(i,j)} & = x(t)^\dagger A^H \\
\frac{\partial s(t)}{\partial \beta(i,j)} & = jx(t)^HBU_j^T
\end{align*}
\]

where \( u_i = [0 \cdots 0, 1, 0 \cdots 0]^T \) is the vector whose \( i \)th element is only unitary.

The elements of \( F_\sigma \) can be calculated using the formula (9). We obtain:

\[
F_{s(i,j)(n,m)} = 2\Re \left[ \sum_{i,j} x(t)^\dagger B^H u_i^T u_j^T B x(t) \right] = 2\Re \left[ Q^{-1}(i,m)u_i^T BXX^H B^H u_j^T \right]
\]

In a similar fashion we calculate:

\[
\begin{align*}
F_{s(i,j)(n,m)} & = 2\Re \left[ Q^{-1}(i,m)(BXX^H B^H) (n,j) \right] \\
F_{s(i,j)(n,m)} & = -2\Im \left[ Q^{-1}(i,m)(BXX^H B^H) (n,j) \right] \\
F_{s(i,j)(n,m)} & = 2\Re \left[ (A^H Q^{-1} A)(l,m)(X X^H) (n,j) \right] \\
F_{s(i,j)(n,m)} & = 2\Re \left[ (A^H Q^{-1} A)(l,m)(X X^H) (n,j) \right] \\
F_{s(i,j)(n,m)} & = -2\Im \left[ (A^H Q^{-1} A)(l,m)(X X^H) (n,j) \right] \\
F_{s(i,j)(n,m)} & = 2\Re \left[ (Q^{-1} A)(l,m)(X X^H B^H) (n,j) \right] \\
F_{s(i,j)(n,m)} & = -2\Im \left[ (Q^{-1} A)(l,m)(X X^H B^H) (n,j) \right] \\
F_{s(i,j)(n,m)} & = 2\Re \left[ (Q^{-1} A)(l,m)(X X^H B^H) (n,j) \right] \\
F_{s(i,j)(n,m)} & = -2\Im \left[ (Q^{-1} A)(l,m)(X X^H B^H) (n,j) \right] \\
F_{s(i,j)(n,m)} & = 2\Re \left[ (Q^{-1} A)(l,m)(BXX^H) (n,j) \right] \\
F_{s(i,j)(n,m)} & = 2\Re \left[ (Q^{-1} A)(l,m)(BXX^H) (n,j) \right]
\end{align*}
\]
Let:

\[ A = 2 \left( B X X^H B^H \right)^T \otimes Q^{-1}. \]

\[ B = 2 \left[ \begin{bmatrix} 0 & I_{q-J} \end{bmatrix} \left[ X X^H \right]^T \left[ \begin{bmatrix} 0 \\ I_{q-J} \end{bmatrix} \right] \right] \otimes \left[ A^H Q^{-1} A \right]. \]

and

\[ C = 2 \left[ \begin{bmatrix} 0 & I_{q-J} \end{bmatrix} X X^H B^H \right]^T \otimes \left[ Q^{-1} A \right]. \]

Note that the pre-multiplication of a matrix by \( \begin{bmatrix} 0 & I_{q-J} \end{bmatrix} \) is obtained by extracting the \((q-J)\) last lines; and the post-multiplication by \( \begin{bmatrix} 0 \\ I_{q-J} \end{bmatrix} \) is obtained by extracting the \((q-J)\) last columns. The combination of these two operations conducts to extract the inferior bloc on the right order \((q-J)\).

We verify from the above expressions of the elements of the Fisher matrix that the FIM can be written using a compact form:

\[ \text{FIM} = \begin{bmatrix} \Re(A) & \Re(C) & -\Im(A) & -\Im(C) \\
\Re(C^T) & \Re(B) & \Im(A^T) & -\Im(B) \\
-\Im(A^T) & \Im(C^T) & \Re(A) & \Re(C) \\
-\Im(C^T) & -\Im(B^T) & \Re(C^T) & \Re(B) \end{bmatrix}. \]

Furthermore, the matrices \( A, B \) and \( C \) have the following properties easy to verify from their definitions:

- \( A \) is Hermitian, and satisfy: \( \Re(A^T) = A, \Im(A^T) = -\Im(A) \)
- \( B \) is Hermitian;
- \( C \) is arbitrary.

**Appendix D: Calculation of the MSE of model 2:**

We demonstrate in this appendix that for the model 2, the MSE of the channel given by the expression (65) is independent of \( A \), when \( XX^H \) and \( Q \) are proportional to the identity (because the power is repeated on the parameterization of \( A \)):

\[ XX^H = \varepsilon I, \quad Q = \sigma^2 I, \quad B = [I \quad B_0] \quad (J'q) \text{ where } B_0 \text{ (J'q-J) is any matrix.} \]
Let us calculate the matrices included in the expression (65). The matrices $V_{aa}$, $V_{bb}$, and $V_{ab}$ are given by the expressions (62), (63) and (64) respectively. The matrices $A$, $B$, and $C$ are expressed by taking into account the values $\Xi$ and $\Omega$:

$$A = \frac{2E}{\sigma^2} \left( (BB^H)^T \otimes I \right)$$  \hspace{1cm} (D.1)

$$B = \frac{2E}{\sigma^2} \left[ I \otimes A^H A \right]$$  \hspace{1cm} (D.2)

and

$$C = \frac{2E}{\sigma^2} \left[ B_0^T \otimes A \right]$$  \hspace{1cm} (D.3)

We deduce:

$$V_{aa} = (A - C B^{-1} C^H)^{-1}$$

$$= (\frac{E}{\sigma^2})^{-1} \left[ (BB^H)^T \otimes \Xi - B_0^T I B_0^T \otimes A (A^H A)^{-1} A^H \right]^{-1}$$  \hspace{1cm} (D.4)

$$V_{bb} = (B - C^H A^{-1} C)^{-1}$$

$$= (\frac{E}{\sigma^2})^{-1} \left[ (I + B_0^T B_0^T) \otimes (\Omega + \Pi^T) - B_0^T B_0^T \otimes \Pi \right]^{-1}$$  \hspace{1cm} (D.5)

$$V_{ab} = -V_{aa} C B^{-1}$$

$$= -V_{aa} (B_0^T \otimes A) (I \otimes (A^H A)^{-1})$$  \hspace{1cm} (D.6)

Thus, the MSE (65) can be written:

$$MSE = MSE_1 + MSE_2 + MSE_3$$
with:

\[ \text{MSE}_1 = \text{tr} \left[ (I_{q,J} \otimes A^H A) V_{ss} \right] = \left( \frac{E}{\sigma^2} \right)^{-1} \text{tr} \left[ (I - B_0^T (B B^H)^{-1} B_0)^{-1} \right] \times J . \]  

\[ \text{MSE}_2 = \text{tr} \left[ (B^T B^T \otimes I_N) V_{ss} \right] = \left( \frac{E}{\sigma^2} \right)^{-1} \left\{ \text{tr} \left[ B^T B^T \right] \times J + J \times (M - J) \right\} , \]  

\[ \text{MSE}_3 = 2 \text{tr} \left[ \left( [I_{q,J}] B^T \otimes A^H \right) V_{ss} \right] = -2 \left( \frac{E}{\sigma^2} \right)^{-1} \text{tr} \left[ B_0^T B_0 \right] \times J . \]  

These imply:

\[ \text{MSE} = \left( \frac{E}{\sigma^2} \right)^{-1} \text{tr} \left[ (I - E_{ss}(BB^T)^{-1} B_s)^{-1} \right] \times J + \left( \frac{E}{\sigma^2} \right)^{-1} \left\{ \text{tr} \left[ B B^T \right] \times J + J \times (M - J) \right\} \]

\[-2 \left( \frac{E}{\sigma^2} \right)^{-1} \text{tr} \left[ B_s B_s \right] \times J . \]  

We deduce that the expression (D.10) of the MSE is independent of A.

![Map of Mathura District](image1)

**Fig. 1**

![Levels of Emigrant in Primary Education](image2)

**Fig. 2**
Fig. 6:

**Table 1:** Mathura District: block-wise mean composite z-score

<table>
<thead>
<tr>
<th>Name Of Blocks</th>
<th>Primary Education</th>
<th>Educational Facilities</th>
<th>Socio-Economic Development</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nanogaon</td>
<td>-0.503</td>
<td>0.094</td>
<td>-0.185</td>
</tr>
<tr>
<td>chhata</td>
<td>0.342</td>
<td>0.157</td>
<td>-0.474</td>
</tr>
<tr>
<td>Chaumuha</td>
<td>-0.168</td>
<td>0.189</td>
<td>-0.263</td>
</tr>
<tr>
<td>Goverdhan</td>
<td>-1.183</td>
<td>-0.315</td>
<td>-0.276</td>
</tr>
<tr>
<td>Mathura</td>
<td>0.991</td>
<td>-0.260</td>
<td>0.256</td>
</tr>
<tr>
<td>Falah</td>
<td>-1.549</td>
<td>-0.146</td>
<td>-0.080</td>
</tr>
<tr>
<td>Nohjheel</td>
<td>0.488</td>
<td>0.230</td>
<td>0.136</td>
</tr>
<tr>
<td>Mant</td>
<td>0.188</td>
<td>-0.008</td>
<td>0.082</td>
</tr>
<tr>
<td>Raya</td>
<td>1.891</td>
<td>0.041</td>
<td>0.821</td>
</tr>
<tr>
<td>Baldeo</td>
<td>-0.557</td>
<td>0.008</td>
<td>0.029</td>
</tr>
</tbody>
</table>

Source: Computed by Authors

**Table 2:** Mathura District: Levels of Enrollment in Primary Education

<table>
<thead>
<tr>
<th>Category</th>
<th>Composite Mean z-score</th>
<th>No.ofblocks</th>
<th>Name of the blocks</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>Above+0.461</td>
<td>3</td>
<td>Mathura, Nohjheel, Raya</td>
</tr>
<tr>
<td>Medium</td>
<td>-0.461 to 0.461</td>
<td>3</td>
<td>Chaumuha, Mant, Chhatta</td>
</tr>
<tr>
<td>Low</td>
<td>Below-0.461</td>
<td>4</td>
<td>Nandgaon, Goverdhan, Bandeo</td>
</tr>
</tbody>
</table>

Source: Computed from BSA, Mathura 2005-06

**Table 3:** Mathura District: Levels of Primary Educational Facilities

<table>
<thead>
<tr>
<th>Category</th>
<th>Composite Mean z-score</th>
<th>No.ofblocks</th>
<th>Name of the blocks</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>Above+0.086</td>
<td>4</td>
<td>Nandgaon, Chhatta, Chaumuha, Nohjheel</td>
</tr>
<tr>
<td>Medium</td>
<td>-0.086 to 0.086</td>
<td>3</td>
<td>Mant, Raya, Baldeo</td>
</tr>
<tr>
<td>Low</td>
<td>Below-0.086</td>
<td>3</td>
<td>Goverdhan, Mathura, Farah</td>
</tr>
</tbody>
</table>

Source: Computed from BSA, Mathura 2005-06 and statistical Handbook 2005-06

**Table 4:** Mathura District: Levels of Socio Economic Development

<table>
<thead>
<tr>
<th>Category</th>
<th>Composite Mean z-score</th>
<th>No.ofblocks</th>
<th>Name of the blocks</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>Above+0.157</td>
<td>2</td>
<td>Mathura, Raya</td>
</tr>
<tr>
<td>Medium</td>
<td>-0.157 to +0.157</td>
<td>4</td>
<td>Nohjheel, Mant, Baldeo, Farah</td>
</tr>
<tr>
<td>Low</td>
<td>Below-0.157</td>
<td>4</td>
<td>Nandgaon, Chhatta, Chaumuha and Goverdhan</td>
</tr>
</tbody>
</table>

Source: Computed from Census of India 2001 and statistical Handbook 2005-06
Table 5: Mathura District: Correlation coefficient ($r$) between Enrollment in Primary education and Indicators (Socio-Economic and Educational facilities)

<table>
<thead>
<tr>
<th>Independent variables (Socio-Economic and Educational facilities)</th>
<th>Dependent Variables (Enrollment in Primary education)</th>
</tr>
</thead>
<tbody>
<tr>
<td>% of S.C. Population ($x_1$)</td>
<td>Total E.P.E (Y1)</td>
</tr>
<tr>
<td>% of total Literate ($x_2$)</td>
<td>-0.069</td>
</tr>
<tr>
<td>% of Urban Population ($x_3$)</td>
<td>0.233</td>
</tr>
<tr>
<td>Average Size of family ($x_4$)</td>
<td>-0.295</td>
</tr>
<tr>
<td>Per capita income ($x_5$)</td>
<td>0.894</td>
</tr>
<tr>
<td>Density of Metalled Roads (Length per 1000 Sq Km) ($x_6$)</td>
<td>0.246</td>
</tr>
<tr>
<td>No of Private primary Schools ($x_7$)</td>
<td>0.710</td>
</tr>
<tr>
<td>Pupil - school Ratio ($x_8$)</td>
<td>-0.166</td>
</tr>
<tr>
<td>No. Of Primary Schools per Lakh Person ($x_{11}$)</td>
<td>0.198</td>
</tr>
<tr>
<td>No. Of upper Primary Schools per Lakh Person ($x_{12}$)</td>
<td>-0.415</td>
</tr>
<tr>
<td>% of Primary school With Water Facilities ($x_{13}$)</td>
<td>-0.125</td>
</tr>
<tr>
<td>% of Primary school With Toilet Facilities ($x_{14}$)</td>
<td>-0.300</td>
</tr>
</tbody>
</table>

Source: Computed by Authors

REFERENCES