Non efficiency and non Gaussianity of a maximum likelihood estimator at high signal-to-noise ratio and finite number of samples

Alexandre Renaux, Philippe Forster, Eric Boyer, Pascal Larzabal

▶ **To cite this version:**

Alexandre Renaux, Philippe Forster, Eric Boyer, Pascal Larzabal. Non efficiency and non Gaussianity of a maximum likelihood estimator at high signal-to-noise ratio and finite number of samples. ICASSP 2004 - 29th IEEE International Conference on Acoustics, Speech and Signal Processing, 2004, Montreal, Canada. 10.1109/icassp.2004.1326209 . inria-00444829

HAL Id: inria-00444829
https://inria.hal.science/inria-00444829

Submitted on 7 Jan 2010

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

NON EFFICIENCY AND NON GAUSSIANITY OF A MAXIMUM LIKELIHOOD ESTIMATOR AT HIGH SIGNAL TO NOISE RATIO AND FINITE NUMBER OF SAMPLES

A. Renaux[(1)], P. Forster[(2)], E. Boyer[(1)] and P. Larzabal[(1)]

[(1)]SATIE, Ecole Normale Supérieure de Cachan, 61 avenue du président Wilson, 94235 CACHAN Cedex FRANCE
[(2)]GEA, IUT de Ville d'Avray, 1 Chemin Desvallières, 92410 VILLE D'AVRAY FRANCE

ABSTRACT

In estimation theory, the asymptotic efficiency of the Maximum Likelihood (ML) method for independent identically distributed observations and when the number T of observations tends to infinity is a well known result. In some scenarii, the number of snapshots may be small making this result unapplicable. In the array processing framework, for Gaussian emitted signals, we fill this lack at high Signal to Noise Ratio (SNR). In this situation, we show that the ML estimation is asymptotically (with respect to SNR) non efficient and non Gaussian.

1. INTRODUCTION

In array processing, the asymptotic performances of the Stochastic Maximum Likelihood (SML) are well known when the number of snapshots T tends to infinity. Indeed, within the context of estimation theory, with independent and identically distributed observations, the asymptotic efficiency and gaussianity of Maximum Likelihood (ML) method, when the number of snapshots T tends to infinity, are well known results [1].

This paper addresses the problem of the SML behaviour for a finite number of samples when the SNR tends to infinity : this is meaning of "asymptotic" in this paper. In this case to our knowledge, there is no general result that gives the maximum likelihood estimator distribution. We show in this paper that the SML estimator is not efficient and not Gaussian when SNR tends to infinity. We also show that at high SNR, Stochastic and Deterministic Maximum Likelihood Criterions (SMLC and DMLC) are equivalent: they yield the same estimates.

This paper is organized as follows. Section 2 presents the signal model. The equivalence between the SML and DML methods is proved in section 3. The asymptotic distribution of SML estimator is established in section 4. To confirm our results, simulations are performed in section 5. Finally, section 6 gives our conclusions.

2. THE MODEL

Let us consider the classical problem of localizing P sources with an array of $M>P$ sensors that receive the signals $\mathbf{y}(t)$:

$$\mathbf{y}(t) = \mathbf{A}(\boldsymbol{\theta}_e)\,\mathbf{x}(t) + \mathbf{n}(t) \quad \text{for } t=1,\dots T, \tag{1}$$

where:

- $\mathbf{A}(\boldsymbol{\theta}) = \left[\mathbf{a}(\theta_1) \;\cdots\; \mathbf{a}(\theta_P)\right]$ is the $M \times P$ matrix of steering vectors;
- $\boldsymbol{\theta} = \left[\theta_1 \;\cdots\; \theta_P\right]^T$ denotes the candidate vector of the P Directions Of Arrival (DOA) whose exact value is $\boldsymbol{\theta}_e$;
- $\mathbf{x}(t)$ is the $P \times 1$ vector of the signals emitted by the P sources;
- $\mathbf{n}(t)$ is the noise. It is modeled as a zero mean complex circular Gaussian vector, spatially and temporally white, with the unknown noise power σ^2 .

In the sequel

$$\mathbf{N} = \left[\mathbf{n}(1), \mathbf{n}(2), \dots, \mathbf{n}(T)\right], \tag{2}$$

and

$$\mathbf{X} = \left[\mathbf{x}(1), \mathbf{x}(2), \dots, \mathbf{x}(T)\right]. \tag{3}$$

3. DMLC AND SMLC

Two models are commonly used for signals sources: the deterministic model, for which the emitted signals $\mathbf{x}(t)$ treated as unknown parameters, and the stochastic model, for which $\mathbf{x}(t)$ is a complex circular Gaussian vector, with zero mean and unknown covariance matrix $\boldsymbol{\Sigma}$, spatially and temporally white.

In the deterministic model case, the DOA are obtained by minimizing of the concentrated criterion:

$$C_{SML}(\boldsymbol{\theta}) = \frac{1}{M-P} Trace\left(\Pi_n(\boldsymbol{\theta})\hat{\mathbf{R}}_{\mathbf{y}}\right), \tag{4}$$

where $\Pi_n(\boldsymbol{\theta})$ denotes the orthogonal projector onto the noise subspace and where the sample covariance matrix of a vector $\boldsymbol{\nu}$ is given by:

$$\hat{\mathbf{R}}_{\mathbf{v}} = \frac{1}{T}\sum_{t=1}^{T} \mathbf{v}(t)\mathbf{v}(t)^H \quad . \tag{5}$$

For the stochastic model case, the DOA are obtained by minimizing of the concentrated criterion:

$$C_{SML}(\boldsymbol{\theta}) = \left| \Pi_s(\boldsymbol{\theta}) \hat{\mathbf{R}}_{\mathbf{y}} \Pi_s(\boldsymbol{\theta}) + C_{DML}(\boldsymbol{\theta}) \Pi_n(\boldsymbol{\theta}) \right|, \tag{6}$$

where $\Pi_s(\boldsymbol{\theta})$ denotes the orthogonal projector onto the signal subspace and $|\mathbf{F}|$ is the determinant of matrix **F**.

Theorem 1
At the first order with respect to **N** the DOA obtained by minimisation of $C_{SML}(\boldsymbol{\theta})$ and $C_{DML}(\boldsymbol{\theta})$ are equal.

Proof
Let $\mathbf{U}_s(\boldsymbol{\theta})$ and $\mathbf{U}_n(\boldsymbol{\theta})$ be the $M \times P$ and $M \times (M-P)$ matrices built with the orthonormal bases of the signal and noise subspaces and set: $\mathbf{U}(\boldsymbol{\theta}) = [\mathbf{U}_s(\boldsymbol{\theta}) \mathbf{U}_n(\boldsymbol{\theta})]$.
From (6) we have

$$C_{SML}(\boldsymbol{\theta}) = \left| \mathbf{U}(\boldsymbol{\theta})^H \left(\Pi_s(\boldsymbol{\theta}) \hat{\mathbf{R}}_{\mathbf{y}} \Pi_s(\boldsymbol{\theta}) + C_{DML}(\boldsymbol{\theta}) \Pi_n(\boldsymbol{\theta}) \right) \mathbf{U}(\boldsymbol{\theta}) \right|. \tag{7}$$

The matrix involved in the determinant of (7) is block diagonal so that $C_{SML}(\boldsymbol{\theta})$ can be also written as:

$$C_{SML}(\boldsymbol{\theta}, \mathbf{n}) = \underbrace{\left| \mathbf{U}_s(\boldsymbol{\theta})^H \hat{\mathbf{R}}_{\mathbf{y}}(\mathbf{n}) \mathbf{U}_s(\boldsymbol{\theta}) \right|}_{\alpha(\boldsymbol{\theta}, \mathbf{n})} C_{DML}(\boldsymbol{\theta}, \mathbf{n})^{M-P}. \tag{8}$$

In this expression, the functional dependence of each term with respect to $\mathbf{n} = vec(\mathbf{N})$ is emphasized. First, notice that $C_{DML}(\boldsymbol{\theta}, \mathbf{n})$ is minimal and null when $(\boldsymbol{\theta}, \mathbf{n}) = (\boldsymbol{\theta}_e, \mathbf{0})$. Therefore, the same result holds for $C_{SML}(\boldsymbol{\theta}, \mathbf{n})$ because $\alpha(\boldsymbol{\theta}, \mathbf{n}) > 0$ (with probability one).
Let us set $\Delta\boldsymbol{\theta} = \boldsymbol{\theta} - \boldsymbol{\theta}_e$. After a series expansion of $\alpha(\boldsymbol{\theta}, \mathbf{n})$ and $C_{DML}(\boldsymbol{\theta}, \mathbf{n})$ around $\boldsymbol{\theta} = \boldsymbol{\theta}_e$ and $\mathbf{n} = \mathbf{0}$, and keeping only the first non vanishing terms, we obtain:

$$\begin{aligned} \alpha(\boldsymbol{\theta}, \mathbf{n}) &= \alpha(\boldsymbol{\theta}_e, \mathbf{0}) + \cdots, \\ C_{DML}(\boldsymbol{\theta}, \mathbf{n}) &= [\Delta\boldsymbol{\theta}^T \mathbf{n}^T] \mathbf{H} [\Delta\boldsymbol{\theta}^T \mathbf{n}^T]^T + \cdots, \end{aligned} \tag{9}$$

where **H** is the Hessian matrix of the C_{DML} criterion and $\alpha(\boldsymbol{\theta}_e, \mathbf{0}) = \left| \mathbf{U}_s(\boldsymbol{\theta}_e)^H \mathbf{A}(\boldsymbol{\theta}_e) \hat{\mathbf{R}}_{\mathbf{x}} \mathbf{A}(\boldsymbol{\theta}_e)^H \mathbf{U}_s(\boldsymbol{\theta}_e) \right| \neq 0$. Therefore criterion C_{SML} is equivalent at high SNR to $\alpha(\boldsymbol{\theta}_e, \mathbf{0}) C_{DML}(\boldsymbol{\theta}, \mathbf{n})^{M-P}$, which proves that minimisation of $C_{SML}(\boldsymbol{\theta})$ and $C_{DML}(\boldsymbol{\theta})$ are equivalent for bearings.

■

4. ASYMPTOTIC DISTRIBUTION OF SML

From now on, concerning signal sources, we are in the stochastic model framework of section 3 and we note $\hat{\boldsymbol{\theta}} = \arg\min C_{SML}(\boldsymbol{\theta})$ the SML estimator. The next theorem establishes the asymptotic distribution of $\hat{\boldsymbol{\theta}}$ and the proof of its non gaussianity.

Theorem 2

When $\sigma \to 0$, $\tilde{\boldsymbol{\theta}} = \frac{1}{\sigma}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_e)$ is asymptotically distributed as $\mathbf{C}\mathbf{y}$ where:

- **y** is a $P \times 1$ standard Gaussian vector;
- **C** is a $P \times P$ random matrix which is independent of vector **y** and whose distribution is detailed below.

The matrix **C** such that

$$\mathbf{C}\mathbf{C}^T = \frac{1}{2T} \{ \mathrm{Re}[\mathbf{H} \odot \mathbf{W}] \}^{-1}, \tag{10}$$

where:

- $\odot$ is the Hadamard product ;
- **W** is complex Wishart distributed with *T* degrees of freedom and parameter matrix the covariance $\boldsymbol{\Sigma}$ of signals sources;
- $\mathbf{H} = \mathbf{D}^H [\mathbf{I} - \mathbf{A}(\mathbf{A}^H \mathbf{A})^{-1} \mathbf{A}^H] \mathbf{D}$, (11)
- $\mathbf{D} = \left[\left(\frac{\partial \mathbf{a}}{\partial \theta} \right)_{\theta_1} \cdots \left(\frac{\partial \mathbf{a}}{\partial \theta} \right)_{\theta_P} \right]$. (12)

Proof of theorem 2

Using theorem 1, at high SNR, both criterions SMLC and DMLC are equivalent for bearing estimation. Therefore consider $\hat{\boldsymbol{\theta}}$ is obtained by minimisation of the DMLC given by (4), although the signals sources are stochastic.
We first study, at high SNR (small σ^2), the conditional distribution $f(\tilde{\boldsymbol{\theta}} | \mathbf{X})$ of $\tilde{\boldsymbol{\theta}}$ given **X** (equation (3)). This asymptotic distribution is Gaussian, with a covariance given by the deterministic Cramer Rao lower bound $\mathbf{B}_D$ [2]

$$\mathbf{B}_D = \frac{1}{2T} \left\{ \mathrm{Re}\left[\mathbf{H} \odot \hat{\mathbf{R}}_{\mathbf{x}}^T \right] \right\}^{-1}. \tag{13}$$

$\hat{\mathbf{R}}_{\mathbf{x}}$ (and thus $\hat{\mathbf{R}}_{\mathbf{x}}^T$) are complex Wishart distributed with *T* degrees of freedom, and parameter matrix $\boldsymbol{\Sigma}$ the covariance of signals sources. Let us set $\mathbf{B}_D = \mathbf{C}\mathbf{C}^T$ the Cholesky factorisation of the $\tilde{\boldsymbol{\theta}}$ covariance conditionally **X**. Therefore, the asymptotic distribution of $\tilde{\boldsymbol{\theta}}$ is the distribution of $\mathbf{C}\mathbf{y}$ where **y** is a complex standard Gaussian random vector. Removing the conditioning on the signals sources, result of theorem 2 is obtained with $\mathbf{W} = \hat{\mathbf{R}}_{\mathbf{x}}^T$.

■

Corollary 1
With the notation of theorem 2, the asymptotic covariance of $\tilde{\boldsymbol{\theta}}$ is given by

$$\mathrm{cov}\left(\tilde{\boldsymbol{\theta}}\right)=\frac{1}{2T}E\left[\left\{\mathrm{Re}\left[\mathbf{H}\odot\mathbf{W}\right]\right\}^{-1}\right]. \tag{14}$$

Corollary 2
For a single source, for which the maximum likelihood estimator amounts to a beamformer, $\tilde{\theta}$ is asymptotically distributed as kS_{2T} where S_{2T} is a Student random variable with $2T$ degrees of freedom and k is given by

$$k^2=\lim_{\sigma\to 0}\frac{1}{\sigma^2}B_S\,, \tag{15}$$

where B_S is the stochastic Cramer Rao lower Bound. The asymptotic variance of $\tilde{\theta}$ is given by

$$\mathrm{var}_{as}\left(\tilde{\theta}\right)=k^2\frac{T}{T-1}\,, \tag{16}$$

and the stochastic maximum likelihood estimator is not efficient.

Proof
For a single source with power Σ, theorem 2 shows that

$$\tilde{\theta}\xrightarrow{\text{a.d.}}c\,y, \tag{17}$$

where y follows a standard normal law (mean value 0 and variance 1) and

$$c=\left(2h\sum_{t=1}^{T}\left|x(t)\right|^2\right)^{-1/2}, \tag{18}$$

where

$$h=\left\|\left(\frac{\partial\mathbf{a}}{\partial\theta}\right)_{\theta_1}\right\|^2-\frac{1}{\left\|\mathbf{a}(\theta_1)\right\|^2}\left|\mathbf{a}(\theta_1)^H\left(\frac{\partial\mathbf{a}}{\partial\theta}\right)_{\theta_1}\right|^2. \tag{19}$$

According to [3], for a single source, the stochastic Cramer Rao lower bound is equal to

$$\begin{aligned}B_S&=\frac{\sigma^2}{2T\Sigma h}\frac{\sigma^2+\Sigma\left\|\mathbf{a}(\theta_1)\right\|^2}{\Sigma\left\|\mathbf{a}(\theta_1)\right\|^2}\\&\approx\frac{\sigma^2}{2T\Sigma h}\ (\text{for }\sigma\to 0).\end{aligned} \tag{20}$$

Furthermore,

$$\frac{2}{\Sigma}\sum_{t=1}^{T}\left|x(t)\right|^2\sim\chi^2_{2T}, \tag{21}$$

where χ^2_{2T} is a chi-square random variable with $2T$ degrees of freedom. When X follows a standard normal law and χ^2_N is independent of X, the ratio $X\Big/\sqrt{\chi^2_N/N}$ follows a Student distribution with N degrees of freedom. Under these conditions, by rising (20) and (21) in (18) and (17) we obtain (15).
Furthermore, $\mathrm{var}(S_N)=N/(N-2)$, and therefore the asymptotic covariance of $\tilde{\theta}$ is given by $k^2T/(T-1)$.

Since $T/(T-1)>1$, the maximum likelihood estimator is not asymptotically efficient in SNR for finite T.

■

5. SIMULATION EXAMPLES

Let us consider an Uniform Linear Array (ULA) of four sensors with half-wavelength spacing with sources located at broadside. We first investigate the case of single source where we have established the estimated DOA's distribution. Secondly, the two sources case is investigated.

5.1 Single source case

In this case, it is easily shown from equation (8) that SMLC and DMLC provide the same estimates whatever SNR. In the case of a source vector, whose norm is independent of the DOA, these criterions amount to a classical beamformer. This is a particular case of theorem 1 for which both criterions are rigorously identical.

Let us consider the DOA estimation of a single source located at zero degree with $T=2$ snapshots. Figure 1 represents the MLE empirical variance (Monte-Carlo simulations are conducted with *10000* independent realizations), the stochastic Cramer Rao lower bound, and the asymptotic theoretical variance given by (16). Here, it is twice the Cramer Rao lower bound. There is a very good agreement between theoretical results and simulations ($T/(T-1)=2$).

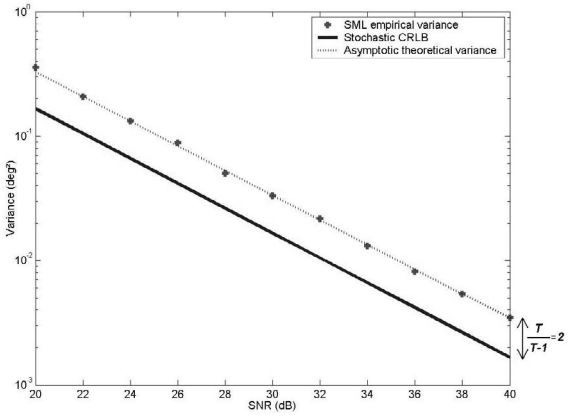

Figure 1: Asymptotic performance of SML estimator (one source)

Figure 2 gives an histogram of ML estimated DOA conducted with *10000* Monte-Carlo simulations, for an SNR of *30*dB. It is in very good adequacy with the theoretical Student law previously established: for comparison, we also reported the probability density function of a Gaussian distribution with the same variance

which is clearly not in adequacy with the distribution of estimates.

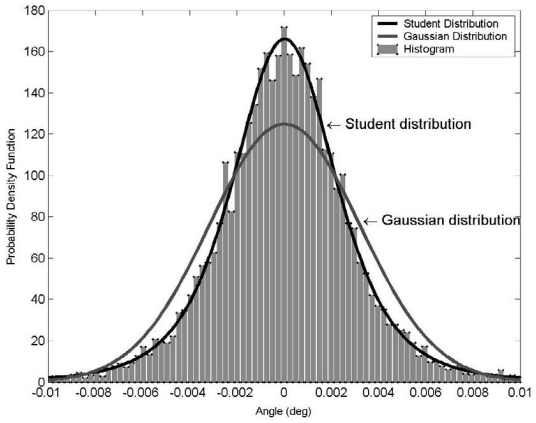

Figure 2: Comparison of estimated bearing distribution and Student distribution with variance given equation (16).

5.2 Two sources case

Let us now consider the case of two sources with same power located at – *7.5* deg and *7.5* deg. The ML DOA estimation is conducted with a Gauss Newton algorithm (initialised at the true value) and *T=3* snapshots. We reported in Figure 3 the evolution of the SML and DML empirical variance (conducted with *10000* Monte Carlo simulations), of the theoretical variance (14), and the stochastic Cramer Rao lower bound. For lack of an analytic expression of (14), the expectation has been estimated by average of *10000* Monte Carlo trials. We note a good agreement between theoretical results and simulations. We can also note the non efficiency of SML at high SNR.

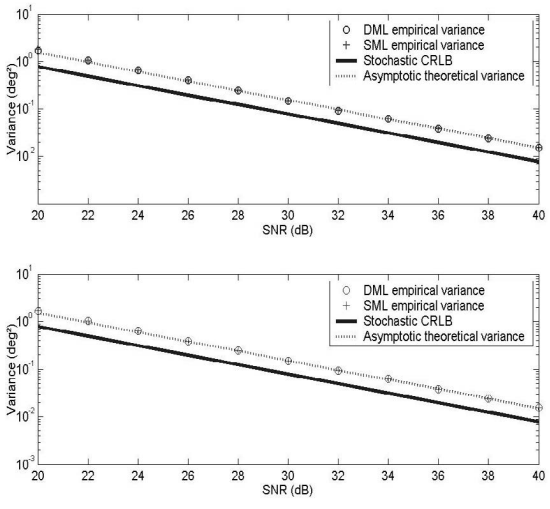

Figure 3 : Asymptotic performance of SML estimator (two sources)

Figure 4 gives the histograms of the estimated DOA corresponding to the previous case. *10000* trials have been performed to compute these histograms (for two sources with the same power and a SNR of *30* dB). Again we note that the probability density function of estimates is not Gaussian.

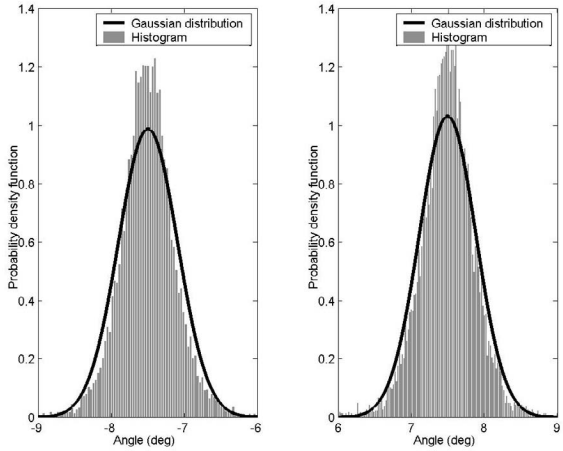

Figure 4: Comparison of estimated bearing distribution and Gaussian distribution with a variance equal to empirical variance.

6. CONCLUSION

In this paper, we have shown that the maximum likelihood estimator is asymptotically (in SNR) non efficient and non Gaussian. Analytical expression of the variance has been derived for the single source case. The analytic extension to the multiple sources case is under investigation.

7. REFERENCES

[1] H. Cramer. *Mathematical methods of statistics*. Princeton University Press, New Jersey, 1946.

[2] P. Stoica and A. Nehorai, "Performance Study of Conditional and Unconditional DOA Estimation," *IEEE Trans. on ASSP*, vol.38, no.10, pp.1783-1795, October1990.

[3] B. Ottersten, M. Viberg, P. Stoica, and A. Nehorai, *Exact and large sample Maximum likelihood techniques for parameter estimation and detection in array processing* pp. 99-151 of *Radar Array Processing*, S. Haykin, J. Litva and T. J. Shepherd, Springler –Verlag, 1993.

[4] A. Eriksson, P. Stoica and T. Söderström, "Second-Order properties of MUSIC and ESPRIT estimates of sinusoidal frequencies in high SNR scenarios," *Radar and Signal Processing, IEE PROCEDINGS-F*, vol.140, no.4, pp.266-272, August 1993.