14th European Signal Processing Conference (EUSIPCO 2006), Florence, Italy, September 4-8, 2006, copyright by EURASIP

STOCHASTIC ML ESTIMATION UNDER MISSPECIFIED NUMBER OF SIGNALS

Pei-Jung Chung

Institute for Digital Communications
Joint Research Institute for Signal & Image Processing
School of Engineering and Electronics, The University of Edinburgh, UK
peijung.chung@gmail.com

ABSTRACT

The maximum likelihood (ML) approach for estimating direction of arrival (DOA) plays an important role in array processing. Its consistency and efficiency have been well established in the literature. A common assumption is that the number of signals is known. In many applications, this information is not available and needs to be estimated. However, the estimated number of signals does not always coincide with the true number of signals. Thus it is crucial to know whether the ML estimator provides any relevant information about DOA parameters under a misspecified number of signals. In the previous study [3], we focused on the deterministic signal model and showed that the ML estimator under a misspecified number of signals converges to a well defined limit. Under mild conditions, the ML estimator converges to the true parameters. In the current work, we extend those results to the stochastic signal model and validate our analysis by simulations.

1. INTRODUCTION

The problem of estimating direction of arrival (DOA) plays a key role in array processing. Among existing methods, the maximum likelihood (ML) approach has the best statistical properties. It is also known to be robust against small sample numbers, signal coherence and closely located sources.

The consistency property of the ML estimator was derived in in [6] [7]. Therein, the number of signals, m, is implicitly assumed to be the *true* one, m_0. However, in many applications, m is usually unknown in advance and needs to be estimated together with the DOA parameters. The commonly used information theoretic approach [8] or the multiple testing procedure [4] for determining the number of signals has their own estimation errors. We may often apply the ML method with an incorrectly chosen number of signals.

Therefore, it is crucial to know *whether the ML estimator provides any useful information about the true parameters even if the assumed number of signals is incorrect*. In the previous study [3], we applied the theory of ML estimation of misspecified model [9] to the deterministic signal model. Our analysis achieved two conclusions. (1) The ML estimator under misspecified numbers of signals converges to a well defined limit. (2) When the signal sources are well separated, the ML estimator converges to the true parameters.

This work extends those results to the stochastic signal model using a similar approach. To find the limiting point of the ML estimator, we minimize the Kullback-Leibler distance between the assumed probability model and the true probability model and derive a criterion similar to the concentrated likelihood function. To get more insight, we assume that the number of signals is much smaller than the number of sensors. In this case, the ML estimator under misspecified number of signals converges to the true parameters.

This paper is outlined as follows. We give a brief description of the signal model in next section. The consistency property of the quasi ML estimator developed by White is presented in section 3. In section 4, we derive a criterion that defines the limiting point for an arbitrarily chosen number of signals. In section 5, we consider a special case in which the number of signals is much smaller than that of the sensors. Then we present and discuss numerical results in section 6. Our concluding remarks are given in section 7.

2. PROBLEM FORMULATION

Consider an array of n sensors receiving m narrow band signals emitted by far-field sources located at $\boldsymbol{\theta}_m = [\ \theta_1,\ldots,\theta_m]^T$. The array output $\boldsymbol{x}(t)$ is described as

$$\boldsymbol{x}(t) = \boldsymbol{H}_m(\boldsymbol{\theta}_m)\boldsymbol{s}_m(t) + \boldsymbol{n}(t), \quad t = 1,\ldots,T \tag{1}$$

where the ith column $\boldsymbol{d}(\theta_i)$ of the matrix

$$\boldsymbol{H}_m(\boldsymbol{\theta}_m) = [\boldsymbol{d}(\theta_1)\cdots\boldsymbol{d}(\theta_i)\cdots\boldsymbol{d}(\theta_m)] \tag{2}$$

represents the steering vector associated with the signal coming from θ_i. The signal vector $\boldsymbol{s}_m(t)$ is considered as a stationary, temporally uncorrelated complex normal process with zero mean and covariance matrix $\boldsymbol{C}_s = E\boldsymbol{s}_m(t)\boldsymbol{s}_m(t)'$ where $(\cdot)'$ denotes the Hermitian transpose. The noise vector $\boldsymbol{n}(t)$ is a spatially and temporally uncorrelated complex normal process with zero mean and covariance matrix $\nu\boldsymbol{I}_n$ where ν is the noise spectral parameter and $\boldsymbol{I}_n$ is an $n\times n$ identity matrix. Thus, the array output $\boldsymbol{x}(t)$ is complex normally distributed with zero mean and covariance matrix

$$\boldsymbol{C}_x = \boldsymbol{H}_m(\boldsymbol{\theta}_m)\boldsymbol{C}_s\boldsymbol{H}_m(\boldsymbol{\theta}_m)' + \nu\boldsymbol{I}_n. \tag{3}$$

Based on the observations $\{\boldsymbol{x}(t)\}_{t=1}^T$ and a pre-specified number of signals, m, the ML estimate $\hat{\boldsymbol{\theta}}_m(T)$ is obtained by minimizing the *negative* concentrated likelihood function [2]

$$l_T(\boldsymbol{\theta}) = \log\det\big(\boldsymbol{P}(\boldsymbol{\theta}_m)\hat{\boldsymbol{C}}_x\boldsymbol{P}(\boldsymbol{\theta}_m) + \hat{\nu}\boldsymbol{P}^{\perp}(\boldsymbol{\theta}_m)\big), \tag{4}$$

$$\hat{\nu} = \frac{1}{n-m}\mathrm{tr}\big(\boldsymbol{P}^{\perp}(\boldsymbol{\theta}_m)\hat{\boldsymbol{C}}_x\big) \tag{5}$$

P.-J. Chung acknowledges support of her position from the Scottish Funding Council and for their support of the Joint Research Institute with the Heriot-Watt University as a component part of the Edinburgh Research Partnership.

where $\boldsymbol{P}(\boldsymbol{\theta}_m)$ represents the projection matrix onto the column space of $\boldsymbol{H}_m(\boldsymbol{\theta}_m)$ and $\boldsymbol{P}^{\perp}(\boldsymbol{\theta}_m) = \boldsymbol{I}_n - \boldsymbol{P}(\boldsymbol{\theta}_m)$. $\hat{\boldsymbol{C}}_x = \frac{1}{T}\sum_{t=1}^{T} \boldsymbol{x}(t)\boldsymbol{x}(t)'$ denotes the sample covariance matrix.

Suppose that the number of signals m is correctly specified, the ML estimator converges to the true parameter $\boldsymbol{\theta}_0$ with increasing sample size [6]. However, as mentioned previously, the number of signals is usually unknown and needs to be estimated via an additional step. Due to estimation errors, the estimated number of signals $\hat{m}$ does not necessarily coincide with the true number of signals m_0. Therefore, it is important to know whether the ML estimate provides any information about $\boldsymbol{\theta}_0$ if one does not assume the correct number of signals.

3. ML ESTIMATES OF MISSPECIFIED MODEL

A misspecified number of signals corresponds to a wrong model order. We treat an incorrectly chosen m as a model mismatch problem. Under a very general framework [9], White defines the ML estimator (MLE) under an incorrectly specified probability model as the quasi ML estimator (QMLE).

More precisely, let $g(\boldsymbol{x})$ and $f(\boldsymbol{x},\boldsymbol{\vartheta})$ denote the true and assumed probability density function of the underlying data $\boldsymbol{x}$, respectively. The QMLE maximizes the log-likelihood function

$$\hat{\boldsymbol{\vartheta}}(T) = \arg\max_{\boldsymbol{\vartheta}} L_T(\boldsymbol{\vartheta}) \tag{6}$$

where

$$L_T(\boldsymbol{\vartheta}) = \frac{1}{T}\sum_{t=1}^{T} \log f(\boldsymbol{x}_t, \boldsymbol{\vartheta}). \tag{7}$$

It is well known that when $f(\boldsymbol{x},\boldsymbol{\vartheta})$ contains the real structure, i.e. $f(\boldsymbol{x},\boldsymbol{\vartheta}_0) = g(\boldsymbol{x})$ for some $\boldsymbol{\vartheta}_0$, the MLE is consistent for $\boldsymbol{\vartheta}_0$ under proper regularity conditions. Without this restriction, Akaike [1] noted that since $L_T(\vartheta)$ is a natural estimator for $E_g[\log f(\boldsymbol{x},\boldsymbol{\vartheta})]$, $\hat{\boldsymbol{\vartheta}}(T)$ is a natural estimator for $\boldsymbol{\vartheta}^*$, the parameter vector that minimizes the Kullback-Leibler information criterion

$$I(g||f) = E_g\left[\log(g(\boldsymbol{x})/f(\boldsymbol{x},\boldsymbol{\vartheta}))\right], \tag{8}$$

where the expectation is taken with respect to the true model $g(\boldsymbol{x})$. Under regularity conditions on $g(\boldsymbol{x})$, $f(\boldsymbol{x},\boldsymbol{\vartheta})$ and $I(g||f)$, White proved the following.

Theorem 2.2 of [9] The QMLE $\hat{\boldsymbol{\vartheta}}(T)$ defined by (6) converges to $\boldsymbol{\vartheta}^*$ as $T \to \infty$ for almost every sequence.

4. CONSISTENCY UNDER MISSPECIFIED NUMBER OF SIGNALS

From the above discussion, we know that consistency still has a meaning for a misspecified probability model. In the following, we shall apply White's results and find the limiting point of the MLE under a misspecified number of signals.

Assuming the number of signals m, each array observation $\boldsymbol{x}(t)$ follows a complex normal distribution $\mathcal{N}^c(\boldsymbol{0},\boldsymbol{C}_x)$ with $\boldsymbol{C}_x$ defined by (3). The log-likelihood function $\log f(\boldsymbol{x})$ (without constant term) corresponding to the *assumed* probability model is as follows

$$\log f(\boldsymbol{x}) = -\left[\ \log\det \boldsymbol{C}_x + \mathrm{tr}\big(\boldsymbol{C}_x^{-1}\boldsymbol{x}(t)\boldsymbol{x}(t)'\big)\ \right]. \tag{9}$$

The *true* probability model $g(\boldsymbol{x})$ corresponding to the correct number of signals, m_0, is given by $\mathcal{N}^c(\boldsymbol{0},\boldsymbol{C}_{x_0})$ where

$$\boldsymbol{C}_{x_0} = \boldsymbol{H}_0\boldsymbol{C}_{s_0}\boldsymbol{H}_0' + \nu_0\boldsymbol{I}_n \tag{10}$$

with the true steering matrix $\boldsymbol{H}_0 = \boldsymbol{H}_{m_0}(\boldsymbol{\theta}_0)$. Since $g(\boldsymbol{x})$ describes the true model, $\boldsymbol{C}_{s_0}$, $\boldsymbol{\theta}_0$, ν_0 are the *true* parameters. They are considered as constant throughout our analysis. Based on **Theorem 2.2** of [9], we obtain a criterion that defines the limiting point of the ML estimator.

Theorem 1 Assume the number of signals to be m. The ML estimator $\boldsymbol{\theta}_m(T)$ converges almost surely to the minimizing point $\boldsymbol{\theta}_m^*$ of the criterion

$$Q(\boldsymbol{\theta}_m) = \log\det\big(\boldsymbol{P}(\boldsymbol{\theta}_m)\boldsymbol{C}_{x_0}\boldsymbol{P}(\boldsymbol{\theta}_m) + \tilde{\nu}\boldsymbol{P}^{\perp}(\boldsymbol{\theta}_m)\big) \tag{11}$$

$$\tilde{\nu} = \frac{1}{n-m}\mathrm{tr}\big(\boldsymbol{P}^{\perp}(\boldsymbol{\theta}_m)\boldsymbol{C}_{x_0}\big) \tag{12}$$

where the true covariance matrix $\boldsymbol{C}_{x_0}$ is given in (10).

Proof: The Kullback-Leibler information criterion (8) can be rewritten as

$$I(g||f) = E_g[\log g(\boldsymbol{x})] - E_g[\log f(\boldsymbol{x})]. \tag{13}$$

Since $g(\boldsymbol{x})$ is fixed, minimizing $I(g||f)$ is equivalent to minimizing the expected negative log-likelihood

$$q = E_g[-\log f(\boldsymbol{x})]\ . \tag{14}$$

Inserting (9) into (14), we obtain

$$\begin{aligned} q &= E_g\left[\ \log\det \boldsymbol{C}_x + \mathrm{tr}(\boldsymbol{C}_x^{-1}\boldsymbol{x}(t)\boldsymbol{x}(t)')\ \right] \\ &= \log\det \boldsymbol{C}_x + \mathrm{tr}(\boldsymbol{C}_x^{-1}\tilde{\boldsymbol{C}}_x) \end{aligned} \tag{15}$$

where $\tilde{\boldsymbol{C}}_x$ is the expected second moment of $\boldsymbol{x}(t)$

$$\tilde{\boldsymbol{C}}_x = E_g[\boldsymbol{x}(t)\boldsymbol{x}(t)'] = \boldsymbol{C}_{x_0}. \tag{16}$$

The criterion q has the same form as the log-likelihood (9). The only difference is that the term $\boldsymbol{x}(t)\boldsymbol{x}(t)'$ is now replaced by its expectation $\boldsymbol{C}_{x_0}$. Applying a similar technique for obtaining the concentrated likelihood function (4), we concentrate (15) with respect to the signal and noise parameters, $\boldsymbol{C}_s$ and ν. Finally we obtain a criterion that depends only on $\boldsymbol{\theta}_m$ as following.

$$Q(\boldsymbol{\theta}_m) = \log\det\big(\boldsymbol{P}(\boldsymbol{\theta}_m)\boldsymbol{C}_{x_0}\boldsymbol{P}(\boldsymbol{\theta}_m) + \tilde{\nu}\boldsymbol{P}^{\perp}(\boldsymbol{\theta}_m)\big) \tag{17}$$

with $\tilde{\nu}$ given by (12). □

Remark: The similarity between criterion $Q(\boldsymbol{\theta}_m)$ and $l_T(\boldsymbol{\theta})$ implies that regardless the assumed number of signals m, the limiting point $\boldsymbol{\theta}_m^*$ minimizes the *expected* negative concentrated likelihood function.

To get more insight into the relation between the true parameters $\boldsymbol{\theta}_0$ and the limiting point $\boldsymbol{\theta}_m^*$, we need the following results.

Result 1 Consider the singular value decomposition (SVD) of the steering matrix $\boldsymbol{H}_m(\boldsymbol{\theta}_m) = \boldsymbol{U}\,\boldsymbol{\Sigma}\,\boldsymbol{V}'$ where $\boldsymbol{U} = [\boldsymbol{U}_1\,\boldsymbol{U}_2]$. $\boldsymbol{U}_1$ consists of the first m columns of $\boldsymbol{U}$ corresponding to the m largest singular values. The criterion $Q(\boldsymbol{\theta}_m)$ in (11) can be expressed as [5]

$$\begin{aligned} Q(\boldsymbol{\theta}_m) &= \log\det\left(\boldsymbol{U}\begin{bmatrix} \boldsymbol{U}_1'\boldsymbol{C}_{x_0}\boldsymbol{U}_1 & \boldsymbol{0} \\ \boldsymbol{0} & \tilde{\nu}\boldsymbol{I}_{n-m}\end{bmatrix}\boldsymbol{U}'\right) \\ &= \log\det\left(\boldsymbol{U}_1'\boldsymbol{C}_{x_0}\boldsymbol{U}_1\right) + (n-m)\log\tilde{\nu}\,. \end{aligned} \qquad (18)$$

Result 2 By the definition of $\boldsymbol{C}_{x_0}$ and $\tilde{\nu}$, and properties of SVD, one can easily verify eqs. (19), (20).

$$\det(\boldsymbol{U}_1'\boldsymbol{C}_{x0}\boldsymbol{U}_1) = \det\left(\boldsymbol{P}(\boldsymbol{\theta}_m)\boldsymbol{H}_0\boldsymbol{C}_{s_0}\boldsymbol{H}_0'\boldsymbol{P}(\boldsymbol{\theta}_m) + \nu_0\boldsymbol{I}_m\right) \quad (19)$$

$$\tilde{\nu} = \frac{1}{m-n}\mathrm{tr}\left(\boldsymbol{P}^{\perp}(\boldsymbol{\theta}_m)\boldsymbol{H}_0\boldsymbol{C}_{s_0}\boldsymbol{H}_0'\boldsymbol{P}^{\perp}(\boldsymbol{\theta}_m)\right) + \nu_0 \qquad (20)$$

Given an assumed number of signals m, we can interpret the first term of (18) as the *signal* part and the second term as the *noise* part.

5. SPECIAL CASE

To simplify our analysis, we assume the number of signals is much smaller than the number of sensors, i.e., $m << n$. In this case, the criterion $Q(\boldsymbol{\theta}_m)$ in (18) is dominated by the noise part $(n-m)\log\tilde{\nu}$.

Case 1: $m = m_0$. Clearly, with the correct number of signals we have $\boldsymbol{\theta}_m^* = \boldsymbol{\theta}_0$. Furthermore,

$$Q(\boldsymbol{\theta}_m^*) = Q(\boldsymbol{\theta}_0) = \sum_{i=1}^{m_0}\log\lambda_i + (n-m_0)\log\nu_0 \qquad (21)$$

where $\lambda_1 \geq \cdots \geq \lambda_{m_0}$ are the m_0 largest eigenvalues of $\boldsymbol{C}_{x_0}$.

Case 2: $m < m_0$, the assumed number of signals is smaller than the true number of signals. According to eq. (20), minimizing $\tilde{\nu}$ is equivalent to minimizing the distance between $\boldsymbol{H}_0$ and $\boldsymbol{P}(\boldsymbol{\theta}_m)\boldsymbol{H}_0$

$$e^2 = \mathrm{tr}\Big(\big(\boldsymbol{I}_n - \boldsymbol{P}(\boldsymbol{\theta}_m)\big)\boldsymbol{H}_0\boldsymbol{C}_{s_0}\boldsymbol{H}_0'\big(\boldsymbol{I}_n - \boldsymbol{P}(\boldsymbol{\theta}_m)\big)\Big). \qquad (22)$$

Since $\mathrm{rank}(\boldsymbol{P}(\boldsymbol{\theta}_m)\boldsymbol{H}_0) < \mathrm{rank}(\boldsymbol{H}_0)$, the best approximation to $\boldsymbol{H}_0$ occurs when $\mathrm{sp}(\boldsymbol{P}(\boldsymbol{\theta}_m)\boldsymbol{H}_0) \subset \mathrm{sp}(\boldsymbol{H}_0)$. For widely separated signals, one can expect that the columns of $\boldsymbol{H}_m(\boldsymbol{\theta}_m)$ coincide with m columns of $\boldsymbol{H}_0$ and the components of $\boldsymbol{\theta}_m^*$ coincide with m components of $\boldsymbol{\theta}_0$.

Furthermore, based on **Result 1** and **Result 2**, the criterion $Q(\boldsymbol{\theta}_m^*)$ can be approximately expressed as

$$Q(\boldsymbol{\theta}_m^*) \approx \sum_{i=1}^{m}\log\lambda_i + \ (n-m)\log(\nu_0+\delta) \qquad (23)$$

where $\delta = e^2/(n-m)$.

Case 3: $m > m_0$, the assumed number of signals is larger than the true number of signals. $\boldsymbol{H}_m(\boldsymbol{\theta}_m)$ has more columns than $\boldsymbol{H}_0$. It is possible that $\mathrm{sp}(\boldsymbol{H}_0) \subset \mathrm{sp}(\boldsymbol{H}_m(\boldsymbol{\theta}_m))$ and $\boldsymbol{P}(\boldsymbol{\theta}_m)\boldsymbol{H}_0 = \boldsymbol{H}_0$. When this happens, e^2 achieves its minimum value zero. For well separated sources, one can expect that m_0 components of $\boldsymbol{\theta}_m^*$ coincide with the components of $\boldsymbol{\theta}_0$. However, the rest $(m-m_0)$ components that correspond to the $(m-m_0)$ columns of $\boldsymbol{H}_m(\boldsymbol{\theta}_m^*)$ are not predictable.

With **Result 1** and **2**, we obtain an approximate expression for $Q(\boldsymbol{\theta}_m^*)$ as follows

$$Q(\boldsymbol{\theta}_m^*) \approx \sum_{i=1}^{m_0}\log\lambda_i + \sum_{i=m_0+1}^{m}\log\nu_0 \ + \ (n-m)\log\nu_0\,. \quad (24)$$

Comparing with eq. (21), we can observe that $Q(\boldsymbol{\theta}_m^*)$ is close to the optimal value $Q(\boldsymbol{\theta}_0)$.

6. SIMULATION

We demonstrate the validity of our analysis by simulation. A uniform linear array of 15 sensors with inter-element spacings of half a wavelength is considered. The narrow band signals are generated by three signal sources of equal strengths located at $[18^\circ\ 36^\circ\ 60^\circ]$. The Signal to Noise Ratio (SNR) is kept at 0 dB. The data consists of $T = 200$ snapshots. The experiment performs 500 trials. In each trial, the DOA parameters are estimated under three assumptions: $m = 2$, $m = m_0 = 3$, $m = 4$. The corresponding estimates are denoted by $\hat{\boldsymbol{\theta}}_2$, $\hat{\boldsymbol{\theta}}_3$, $\hat{\boldsymbol{\theta}}_4$, respectively.

Fig.1 displays the histogram of $\hat{\boldsymbol{\theta}}_2$'s components. In this case, $m < m_0$, the number of signals is smaller than the true number of signals. $\hat{\boldsymbol{\theta}}_2$ has less components than $\boldsymbol{\theta}_0$. The components of $\hat{\boldsymbol{\theta}}_2$ are centered at the true DOAs. The ML estimates are very close to the true parameters.

For comparison, we present results obtained from the true model case, $m = m_0 = 3$ in fig.2. As expected, $\hat{\boldsymbol{\theta}}_3$'s components are distributed around the true parameters.

When the number of signals is larger than the true one, i.e. $m > m_0$, there are $(m-m_0)$ redundant steering vectors. For $m = 4, m_0 = 3$, we have one redundant vector. The empirical distribution of $\hat{\boldsymbol{\theta}}_4$ is illustrated in fig. 3. Similar to figs. 1 and 2, three peaks are positioned at the true parameter values. However, the estimates corresponding to the redundant steering vector are distributed over the entire parameter range. It is not surprising because the redundant vector only needs to lie in a three dimensional subspace spanned by the columns of $\boldsymbol{H}_0$. There is no other restriction on it.

In summary, we observe that ML estimates give us relevant information about the true DOA parameters for correctly and incorrectly chosen m.

7. CONCLUSION

We investigate the consistency property of stochastic ML estimates when the number of signals is incorrectly specified. From the theory of ML estimation under misspecified probability model, we know that the quasi ML estimator converges to a well defined point that minimizes the Kullback-Leibler criterion. Based on this result, we derived a criterion that defines the limiting point. Furthermore, we considered a special case in which the assumed number of signals is much smaller than the number of sensors. In this case, the ML estimator converges to true DOA parameters. Simulations showed that ML estimates give us relevant information about the true parameters even with a misspecified number of signals.

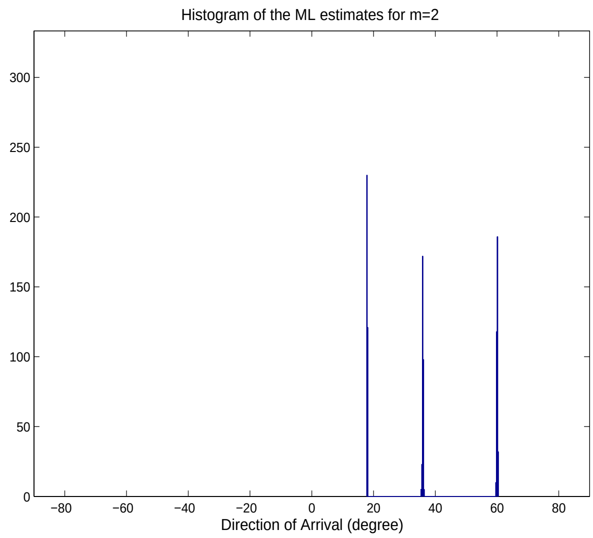

Fig. 1. Histogram of the ML estimates $\hat{\boldsymbol{\theta}}_2$ for the assumed number of signals $m = 2$. The true DOA parameter $\boldsymbol{\theta}_3 = [18^\circ\ 36^\circ\ 60^\circ]$, SNR $= [0\ 0\ 0]$ dB.

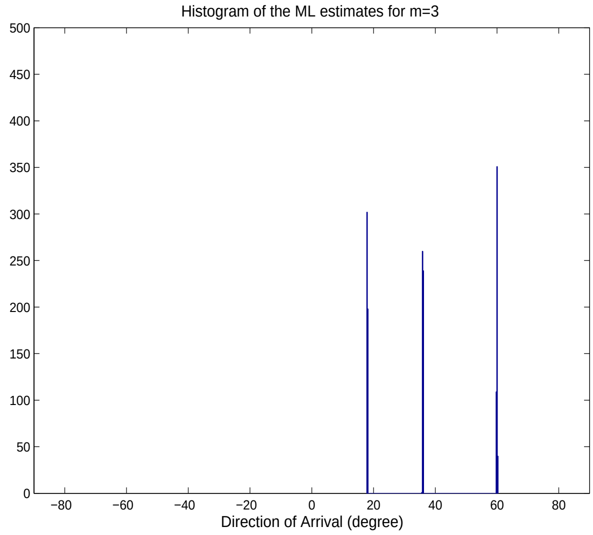

Fig. 2. Histogram of the components of $\hat{\boldsymbol{\theta}}_3$ for the assumed number of signals $m = 3$. The true DOA parameter $\boldsymbol{\theta}_3 = [18^\circ\ 36^\circ\ 60^\circ]$, SNR $=$ $[0\ 0\ 0]$ dB.

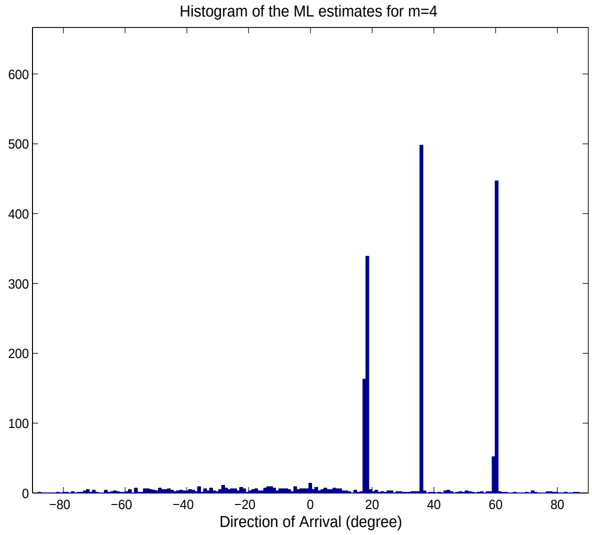

Fig. 3. Histogram of the components of $\hat{\boldsymbol{\theta}}_4$ for the assumed number of signals $m = 4$. The true DOA parameter $\boldsymbol{\theta}_3 = [18^\circ\ 36^\circ\ 60^\circ]$, SNR $=$ $[0\ 0\ 0]$ dB.

8. REFERENCES

[1] H. Akaike. Information theory and an extension to likelihood principle. In B. N. Petrov and F. Csáki, editors, *Proc. of the Second Symposium of Information Theory*, Budapest, 1973.

[2] Johann F. Böhme. Statistical array signal processing of measured sonar and seismic data. In *Proc. SPIE 2563 Advanced Signal Processing Algorithms*, pages 2–20, San Diego, Jul 1995.

[3] Pei-Jung Chung. ML estimation under misspecified number of signals. In *Proc. of the 39th Asilomar Conference on Signals, Systems, and Computers*, Pacific Grove, CA, USA, Oct 30 - Nov 2, 2005.

[4] Pei-Jung Chung, Johann F. Böhme, Alfred O. Hero, and Christoph F. Mecklenbräuker. Signal detection using a multiple hypothesis test. In *Proc. Third IEEE Sensor Array and Multichannel Signal Processing Workshop*, Barcelona, Spain, July 18-21, 2004.

[5] D. Kraus. *Approximative Maximum-Likelihood-Schätzung und verwandte Verfahren zur Ortung und Signalschätzung mit Sensorgruppen*. Dr.–Ing. Dissertation, Faculty of Electrical Engineering, Ruhr–Universität Bochum, Shaker Verlag, Aachen, 1993.

[6] Petre Stoica and A.Nehorai. MUSIC, maximum likelihood and Cramer–Rao bound. *IEEE Trans. Acoustics, Speech, and Signal Processing*, 37(5):720–741, May 1989.

[7] Petre Stoica and Arye Nehorai. Performance study of conditional and unconditional direction–of–arrival estimation. *IEEE Trans. ASSP*, 38(10):1783–1795, October 1990.

[8] M. Wax and I. Ziskind. Detection of the number of coherent signals by the mdl principle. *IEEE ASSP*, 37(8):1190–1196, August 1989.

[9] Halbert White. Maximum likelihoood estimation of misspecified models. *Econometrica*, 50(1):1–25, Jan 1982.