Conditional versus Unconditional Direction of Arrival Estimation – Key Notes

Saif K. Mohammed

We briefly introduce the *conditional* and *unconditional* estimation models for estimating the direction of arrivals (DoA) in array processing. The key differences are explained and the performance of the maximum likelihood estimate under these two models is discussed. In depth discussion and references can be found in [1]. It is also noted that these two models are not just limited to DoA estimation, but are more general [7].

I. System Model

We consider the problem of estimating the direction of arrivals of signals transmitted by emitters in the far field, using the observations at an array with m sensor elements. Let the array response to a signal emitted from an angle θ (this could be a single vertical or azimuthal angle, or could be a vector of both, but here we consider this to be a scalar variable) be denoted by the m-dimensional complex vector $\boldsymbol{a}(\theta)$. We further assume that the number of emitters is known to be d. In general, this can be estimated from the observations, and will be dealt with in subsequent lectures. Let the arrival angle from the k-th emitter be denoted by θ_k, and $\boldsymbol{\theta} \stackrel{\Delta}{=} (\theta_1\, \theta_2\, \cdots\, \theta_d)^T$ denote the vector of all the arrival angles. Further, let $\boldsymbol{A}(\boldsymbol{\theta}) = (\boldsymbol{a}(\theta_1)\, \boldsymbol{a}(\theta_2)\, \cdots\, \boldsymbol{a}(\theta_d))$ denote the complex $m \times d$ matrix with $\boldsymbol{a}(\theta_k)$ as its k-th column. For unique identifiability, we assume that for any full rank $d \times d$ matrices $\boldsymbol{T}$ and $\boldsymbol{U}$,

$$\boldsymbol{A}(\boldsymbol{\theta})\boldsymbol{T} = \boldsymbol{A}(\boldsymbol{\Phi})\boldsymbol{U} \implies \boldsymbol{\theta} = \boldsymbol{\Phi}. \tag{1}$$

Let $x_k(t)$ be the signal emitted from the k-th emitter at time t. By $\boldsymbol{x}(t) = (x_1(t)\, x_2(t)\, \cdots\, x_d(t))^T$ we denote the vector of signals emitted by the d-emitters at time t. We assume that the time indices are discrete, and exactly N observations are made, i.e., $t = 1, 2, \ldots, N$. Further let the addtive noise at the i-th sensor be modeled by a zero mean temporally white stationary Gaussian random process $w_i(t)$. We further assume that the noise processes across the different sensors are independent and identical in distribution. Let $\boldsymbol{w}(t) = (w_1(t)\, w_2(t)\, \ldots\, w_m(t))^T$ represent the vector of additive nosie samples at time t. Further, let $y_m(t)$ denote the observation at the m-th sensor at time t, and $\boldsymbol{y}(t) = (y_1(t)\, y_2(t)\, \ldots\, y_m(t))^T$ denote the observation vector at the sensor array at time t. At the sensor array, the emitter signals superimpose on each other and therefore

$$\boldsymbol{y}(t) = \boldsymbol{A}(\boldsymbol{\theta})\boldsymbol{x}(t) + \boldsymbol{w}(t)\ ,\ t = 1, 2, \ldots, N. \tag{2}$$

The noise vector satisfies

$$\begin{aligned} \mathbb{E}(\boldsymbol{w}(t)\boldsymbol{w}(s)^H) &= \sigma\delta(t,s)\boldsymbol{I}\ ,\ t,s = 1,2,\ldots,N \\ \mathbb{E}(\boldsymbol{w}(t)\boldsymbol{w}(s)^T) &= \boldsymbol{0}\ ,\ t,s = 1,2,\ldots,N \\ \mathbb{E}(\boldsymbol{w}(t)) &= \boldsymbol{0}\ ,\ t = 1,2,\ldots,N. \end{aligned} \tag{3}$$

We further assume the emitter signal sequence $\{\boldsymbol{x}(1)\, , \boldsymbol{x}(2)\, , \ldots ,\, \boldsymbol{x}(N)\}$ to be a sample from a temporally uncorrelated random process with zero mean and non-singular covariance matrix, i.e.

$$\begin{aligned} \mathbb{E}(\boldsymbol{x}(t)\boldsymbol{x}(s)^H) &= \delta(t,s)\boldsymbol{S}\ ,\ t,s = 1,2,\ldots,N\ ,\ (\det(S) \neq 0) \\ \mathbb{E}(\boldsymbol{x}(t)\boldsymbol{x}(s)^T) &= \boldsymbol{0}\ ,\ t,s = 1,2,\ldots,N. \end{aligned} \tag{4}$$

This implies (by the law of large numbers) that , $\boldsymbol{x}(t)$ is second-order ergodic, i.e.

$$\lim_{N\to\infty} \frac{1}{N} \sum_{t=1}^{M} \boldsymbol{x}(t)\boldsymbol{x}(t)^H) \quad = \quad \boldsymbol{S}. \tag{5}$$

It is also assumed that the signals $\boldsymbol{x}(t)$ and the noise $\boldsymbol{w}(s)$ are independent for all t and s.

II. Conditional and Unconditional estimation models

The primary objective in array processing is the estimation of the parameter $\boldsymbol{\theta}$ based upon the array observations $\boldsymbol{Y}_N \triangleq (\boldsymbol{y}(1)\,,\boldsymbol{y}(2)\,,\ldots\,,\boldsymbol{y}(N))$. However, there are two different models of estimating $\boldsymbol{\theta}$, which depend mainly on our choice of unknown parameters which we want to estimate.

From (2) it is clear that, the unknown parameters are the arrival angles $\boldsymbol{\theta}$, the emitter signal sequence $\{\boldsymbol{x}(1)\,,\boldsymbol{x}(2)\,,\ldots,\,\boldsymbol{x}(N)\}$ and the noise variance σ.

Many applications are only interested in estimating the direction of arrival (DoA), and *not* the actual emitter signals. For such applications, the signal sequence can be assumed to be a random sample of a zero mean temporally uncorrelated Gaussian random process whose stochastic properties are dependent only on the covariance matrix $\boldsymbol{S}$. Therefore, the unknown parameters to be estimated for such an application would be

$$\boldsymbol{\eta}_u = (\boldsymbol{\theta}\,,\ \boldsymbol{S}\,,\ \sigma). \tag{6}$$

We shall refer to this model as the "Uncondtional" model of estimation[1]. Note that the number of unknown real-valued parameters are $d^2 + d + 1$.

In other applications, it may be required to estimate the emitter signals in addition to the direction of arrivals. For such applications, the emitter signal sequence would be assumed to be *deterministic* yet known, that is, for all possible realizations of the noise process, the emitter signal sequence is fixed. Further, the emitter sequence is simply required to be second-order ergodic (i.e., satisfying (5)) . The emitter waveforms need not be the output of a Gaussian random process, and can be a realization of some stochastic process with an arbitrary distribution. The unknown parameters to be estimated are

$$\boldsymbol{\eta}_c = (\boldsymbol{\theta}\,,\ \boldsymbol{x}(1), \boldsymbol{x}(2), \ldots, \boldsymbol{x}(N)\,,\ \sigma). \tag{7}$$

We shall refer to this model as the "Conditional" model of estimation [2]. In the Conditional model, it is further noted that the number of unknown real-valued parameters is $d(2N+1)+1$, which *increases* with N. From the above definitions, it appears that, since the Conditional model makes no assumptions on the distribution of the emitter signals, Conditional DoA estimates may have a lesser error variance compared to the Unconditional DoA estimates when the emitter signals have arbitrary distribution. However, we shall see later that the exact opposite is true asymptotically (as $N \to \infty$).

III. Maximum likelihood DoA estimation : Unconditional model

In this section we discuss the maximum likelihood estimation of the DoA angles under the Unconditional model assumption (UMA). We derive expressions for the asymptotic covariance matrix of the ML estimate, and also the Cramer-Rao lower bound (CRLB).

Under UMA, the joint density function of the observations ($\boldsymbol{Y}_N$) for a given unknown but deterministic value of the parameters $\boldsymbol{\eta}_u = (\boldsymbol{\theta}\,,\ \boldsymbol{S}\,,\ \sigma)$ is given by

$$p_u(\boldsymbol{Y}_N; \boldsymbol{\theta}, \boldsymbol{S}, \sigma) = \left(\frac{1}{\pi^m \det(\boldsymbol{R})}\right)^N e^{-\mathrm{tr}(\boldsymbol{R}^{-1} \sum_{t=1}^{N} \boldsymbol{y}(t)\boldsymbol{y}(t)^H)} \tag{8}$$

[1]In some texts, this model has also been referred to as the "Stochastic" model.

[2]In some texts, this model has also been referred to as the "Deterministic" model.

where

$$\boldsymbol{R} \triangleq \mathbb{E}(\boldsymbol{y}(t)\boldsymbol{y}(t)^H) = \boldsymbol{A}(\boldsymbol{\theta})\boldsymbol{S}\boldsymbol{A}(\boldsymbol{\theta})^H + \sigma \boldsymbol{I} \tag{9}$$

is the covariance of the observed signal. The maximum likelihood estimate (MLE) of the parameters $\boldsymbol{\eta}_u = (\boldsymbol{\theta}\,,\, \boldsymbol{S}\,,\, \sigma)$ is then given by

$$\begin{aligned}
\widehat{\boldsymbol{\eta}}_{uml} &= \arg \max_{\boldsymbol{\theta}, \boldsymbol{S} \in \mathcal{S}, \sigma > 0} p_u(\boldsymbol{Y}_N; \boldsymbol{\theta}, \boldsymbol{S}, \sigma) \\
&= \arg \min_{\boldsymbol{\theta}, \boldsymbol{S} \in \mathcal{S}, \sigma > 0} \text{logdet}\Big(\boldsymbol{A}(\boldsymbol{\theta})\boldsymbol{S}\boldsymbol{A}(\boldsymbol{\theta})^H + \sigma \boldsymbol{I}\Big) + \text{tr}\Big((\boldsymbol{A}(\boldsymbol{\theta})\boldsymbol{S}\boldsymbol{A}(\boldsymbol{\theta})^H + \sigma \boldsymbol{I})^{-1}\widehat{\boldsymbol{R}}\Big)
\end{aligned} \tag{10}$$

where $\mathcal{S}$ is the set of all $d \times d$ complex Hermitian positive definete matrices and

$$\widehat{\boldsymbol{R}} \triangleq \frac{1}{N}\sum_{t=1}^{N} \boldsymbol{y}(t)\boldsymbol{y}(t)^H = \frac{1}{N}\boldsymbol{Y}_N \boldsymbol{Y}_N^H \tag{11}$$

is the sampled autocovariance matrix of the observed signal.

The likelihood function can be concentrated w.r.t. the DoA parameters [2]–[4]. With this approach, the MLE is given by

$$\begin{aligned}
\widehat{\sigma}_{uml} &= \frac{\text{tr}(\boldsymbol{P}^{\perp}_{\boldsymbol{A}(\widehat{\boldsymbol{\theta}}_{uml})}\widehat{\boldsymbol{R}})}{m-d} \\
\widehat{\boldsymbol{S}}_{uml} &= \boldsymbol{A}(\widehat{\boldsymbol{\theta}}_{uml})^{\dagger}\Big(\widehat{\boldsymbol{R}} - \widehat{\sigma}_{uml}\boldsymbol{I}\Big)\boldsymbol{A}(\widehat{\boldsymbol{\theta}}_{uml})^{\dagger^H}
\end{aligned} \tag{12}$$

where $\boldsymbol{P}^{\perp}_{\boldsymbol{A}} \triangleq (\boldsymbol{I} - \boldsymbol{A}(\boldsymbol{A}^H\boldsymbol{A})^{-1}\boldsymbol{A}^H)$ is the projector matrix for the subspace of vectors orthogonal to the range space of $\boldsymbol{A}$, and $\boldsymbol{A}^{\dagger} \triangleq (\boldsymbol{A}^H\boldsymbol{A})^{-1}\boldsymbol{A}^H$ is the pseudo-inverse of $\boldsymbol{A}$. Further, $\widehat{\boldsymbol{\theta}}_{uml}$ is the MLE of the DoA under UMA, and is given by

$$\widehat{\boldsymbol{\theta}}_{uml} = \arg \min_{\boldsymbol{\theta}} V_{uml}(\boldsymbol{\theta}) \tag{13}$$

where

$$\begin{aligned}
V_{uml}(\boldsymbol{\theta}) &\triangleq \log\det\Big(\boldsymbol{A}(\boldsymbol{\theta})\widehat{\boldsymbol{S}}(\boldsymbol{\theta})\boldsymbol{A}(\boldsymbol{\theta})^H + \widehat{\sigma}(\boldsymbol{\theta})\boldsymbol{I}\Big) \\
\widehat{\boldsymbol{S}}(\boldsymbol{\theta}) &\triangleq \boldsymbol{A}(\boldsymbol{\theta})^{\dagger}\Big(\widehat{\boldsymbol{R}} - \widehat{\sigma}(\boldsymbol{\theta})\boldsymbol{I}\Big)\boldsymbol{A}(\boldsymbol{\theta})^{\dagger^H} \\
\widehat{\sigma}(\boldsymbol{\theta}) &\triangleq \frac{\text{tr}(\boldsymbol{P}^{\perp}_{\boldsymbol{A}(\boldsymbol{\theta})}\widehat{\boldsymbol{R}})}{m-d}
\end{aligned} \tag{14}$$

Asymptotic properties of the ML estimate under UMA It can be shown that the ML estimate $\widehat{\boldsymbol{\eta}}_{uml}$ is a consistent estimate, and is asymptotically efficient [5]. Therefore if the true value of the DoA parameters is $\boldsymbol{\theta}_0$, then the normalized ML estimation error for the DoA, i.e., $\sqrt{N}(\widehat{\boldsymbol{\theta}}_{uml} - \boldsymbol{\theta}_0)$ is asymptotically Gaussian distributed with zero mean and a covariance matrix $\boldsymbol{C}_{uml}$ which is N times the CRLB for the estimation of DoA under UMA. (*Note that the basic principles of asymptotic analysis was covered in more detail in Lecture 2*).

For the sake of completeness we present the CRLB for the estimation of DoA in the following. If $\widehat{\boldsymbol{\theta}}$ is an unbiased estimate of the DoA i.e., $\mathbb{E}(\widehat{\boldsymbol{\theta}}) = \boldsymbol{\theta}_0$ then the CRLB on the estimation error covariance is given by

$$\begin{aligned}
\mathbb{E}\Big((\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^T\Big) &\succcurlyeq \boldsymbol{B}_u\,, \text{ where} \\
\boldsymbol{B}_u &= \frac{\sigma}{2N}\Big[\text{Re}\Big\{(\boldsymbol{D}(\boldsymbol{\theta}_0)^H\boldsymbol{P}^{\perp}_{\boldsymbol{A}(\boldsymbol{\theta}_0)}\boldsymbol{D}(\boldsymbol{\theta}_0)) \odot (\boldsymbol{S}\boldsymbol{A}(\boldsymbol{\theta}_0)^H\boldsymbol{R}(\boldsymbol{\theta}_0)^{-1}A(\boldsymbol{\theta}_0)\boldsymbol{S})^T\Big\}\Big]^{-1}
\end{aligned} \tag{15}$$

where

$$\begin{aligned}\boldsymbol{D}(\boldsymbol{\theta}) &\triangleq \left(\frac{\partial \boldsymbol{a}(\theta)}{\partial \theta}\Big|_{\theta=\theta_1}, \frac{\partial \boldsymbol{a}(\theta)}{\partial \theta}\Big|_{\theta=\theta_2}, \ldots, \frac{\partial \boldsymbol{a}(\theta)}{\partial \theta}\Big|_{\theta=\theta_d}\right) \\ \boldsymbol{R}(\boldsymbol{\theta}) &\triangleq \boldsymbol{A}(\boldsymbol{\theta})\boldsymbol{S}\boldsymbol{A}(\boldsymbol{\theta})^H + \sigma \boldsymbol{I}. \end{aligned} \tag{16}$$

For a detailed proof please refer to [5].

IV. Maximum Likehood DoA estimation : Conditional model

As discussed previously, under the conditional model assumption (CMA), the emitter signals are assumed to be deterministic yet unknown. There is no assumption made on the distribution of the emitter signals. The only assumption we make is that, it is second-order ergodic.

The probability density function (p.d.f.) of the observations for a given value of parameters $\boldsymbol{\eta}_c = (\boldsymbol{\theta}, \boldsymbol{x}(1), \ldots, \boldsymbol{x}(N), \sigma)$ is given by

$$p_c(\boldsymbol{Y}_N; \boldsymbol{\theta}, \boldsymbol{x}(1), \ldots, \boldsymbol{x}(N), \sigma) = \frac{1}{(\pi\sigma)^{mN}} e^{-\frac{\sum_{t=1}^{N} \|\boldsymbol{y}(t) - \boldsymbol{A}(\boldsymbol{\theta})\boldsymbol{x}(t)\|^2}{\sigma}}. \tag{17}$$

The maximum likelihood estimate under CMA is given by

$$\begin{aligned}(\widehat{\boldsymbol{\theta}}_{cml}, \widehat{\boldsymbol{x}}(1)_{cml}, \ldots, \widehat{\boldsymbol{x}}(N)_{cml}, \widehat{\sigma}_{cml}) &= \arg \max_{\boldsymbol{\theta}, \boldsymbol{x}(1), \ldots, \boldsymbol{x}(N), \sigma>0} p_c(\boldsymbol{Y}_N; \boldsymbol{\theta}, \boldsymbol{x}(1), \ldots, \boldsymbol{x}(N), \sigma) \\ &= \arg \min_{\boldsymbol{\theta}, \boldsymbol{x}(1), \ldots, \boldsymbol{x}(N), \sigma>0} Nm\log(\pi\sigma) + \frac{\|\boldsymbol{Y}_N - \boldsymbol{A}(\boldsymbol{\theta})\boldsymbol{X}_N\|_F^2}{\sigma} \\ &= \arg \min_{\boldsymbol{\theta}, \boldsymbol{x}(1), \ldots, \boldsymbol{x}(N), \sigma>0} l(\boldsymbol{\theta}, \boldsymbol{X}_N, \sigma) \end{aligned} \tag{18}$$

where $\boldsymbol{X}_N \triangleq (\boldsymbol{x}(1), \ldots, \boldsymbol{x}(N))$ and $l(\boldsymbol{\theta}, \boldsymbol{X}_N, \sigma) \triangleq Nm\log(\pi\sigma) + \frac{\|\boldsymbol{Y}_N - \boldsymbol{A}(\boldsymbol{\theta})\boldsymbol{X}_N\|_F^2}{\sigma}$ is the negative log-likelihood function.

This problem is easily separable and can be concentrated w.r.t. $\boldsymbol{\theta}$. Note that the concept of "concentration" of a likelihood function is important and widely used. In many applications, there are many more unknown nuissance parameters than the unknown parameters of interest. The concentration of the likelihood w.r.t. the parameters of interest is therefore of utmost importance from the point of view of reducing the dimensionality of searching for optimal estimates of the parameters of interest. Please refer to [8] for a discussion on various methods for concentration of likelihood functions. Also, more discussion and techniques on concentrating Cramer-Rao bound w.r.t. parameters of interest can be found in [9].

Coming back to the problem of concentrating the likelihood function w.r.t. $\boldsymbol{\theta}$, firstly we note that for any given $\boldsymbol{\theta}$ and σ, the optimal value of $\boldsymbol{X}_N$ is given by

$$\begin{aligned}\boldsymbol{X}_N(\boldsymbol{\theta}, \sigma) &= \arg \min_{\boldsymbol{x}(1), \ldots, \boldsymbol{x}(N)} l(\boldsymbol{\theta}, \boldsymbol{X}_N, \sigma) \\ &= \boldsymbol{A}(\boldsymbol{\theta})^{\dagger} \boldsymbol{Y}_N. \end{aligned} \tag{19}$$

Therefore in the minimization problem in (18), using (19) we can substitute for the optimum value of $\boldsymbol{X}_N$ for a given $(\boldsymbol{\theta}, \sigma)$ resulting in

$$\begin{aligned}(\widehat{\boldsymbol{\theta}}_{cml}, \widehat{\sigma}_{cml}) &= \arg \min_{\boldsymbol{\theta}, \sigma>0} l(\boldsymbol{\theta}, \boldsymbol{X}_N(\boldsymbol{\theta}, \sigma), \sigma) \\ &= \arg \min_{\boldsymbol{\theta}, \sigma>0} m\log(\pi\sigma) + \frac{\mathrm{tr}(\boldsymbol{P}^{\perp}_{\boldsymbol{A}(\boldsymbol{\theta})} \widehat{\boldsymbol{R}})}{\sigma} \\ &= \arg \min_{\boldsymbol{\theta}, \sigma>0} l_1(\boldsymbol{\theta}, \sigma) \end{aligned} \tag{20}$$

where $l_1(\boldsymbol{\theta}, \sigma)$ is the negative log-likelihood function concentrated w.r.t. $(\boldsymbol{\theta}, \sigma)$.

We can further concentrate $l_1(\boldsymbol{\theta}, \sigma)$ w.r.t. $\boldsymbol{\theta}$ as follows. Note that for a given $\boldsymbol{\theta}$, the optimal value of σ is given by

$$\begin{aligned} \sigma(\boldsymbol{\theta}) &= \arg\min_{\sigma>0} l_1(\boldsymbol{\theta}, \sigma) \\ &= \frac{\text{tr}(\boldsymbol{P}^{\perp}_{\boldsymbol{A}(\boldsymbol{\theta})}\widehat{\boldsymbol{R}})}{m}. \end{aligned} \tag{21}$$

Using this, $l_1(\boldsymbol{\theta}, \sigma)$ can be further concentrated w.r.t. $\boldsymbol{\theta}$, and therefore

$$\begin{aligned} \widehat{\boldsymbol{\theta}}_{cml} &= \arg\min_{\boldsymbol{\theta}} l_1(\boldsymbol{\theta}, \sigma(\boldsymbol{\theta})) \\ &= \arg\min_{\boldsymbol{\theta}} \text{tr}(\boldsymbol{P}^{\perp}_{\boldsymbol{A}(\boldsymbol{\theta})}\widehat{\boldsymbol{R}}) \\ &= \arg\min_{\boldsymbol{\theta}} V^N_{cml}(\boldsymbol{\theta}) \end{aligned} \tag{22}$$

where $V^N_{cml}(\boldsymbol{\theta}) \triangleq \text{tr}(\boldsymbol{P}^{\perp}_{\boldsymbol{A}(\boldsymbol{\theta})}\widehat{\boldsymbol{R}})$.

From (19) and (21), it follows that the ML estimates for the emitter signal and the noise variance are given by

$$\begin{aligned} \widehat{\sigma}_{cml} &= \frac{\text{tr}(\boldsymbol{P}^{\perp}_{\boldsymbol{A}(\widehat{\boldsymbol{\theta}}_{cml})}\widehat{\boldsymbol{R}})}{m} \\ \widehat{\boldsymbol{X}_{N}}_{cml} &= \boldsymbol{A}(\widehat{\boldsymbol{\theta}}_{cml})^{\dagger}\boldsymbol{Y}_N. \end{aligned} \tag{23}$$

Asymptotic properties of the ML estimates under CMA

It can be shown that the estimation for the DoA is consistent under CMA, since the function $V^N_{cml}(\boldsymbol{\theta})$ converges uniformly w.p. 1 to $V_{cml}(\boldsymbol{\theta})$ as $N \to \infty$, where the limiting function $V_{cml}(\boldsymbol{\theta})$ is given by

$$V_{cml}(\boldsymbol{\theta}) = \text{tr}(\boldsymbol{P}^{\perp}_{\boldsymbol{A}(\boldsymbol{\theta})}(\boldsymbol{A}(\boldsymbol{\theta})\boldsymbol{S}\boldsymbol{A}(\boldsymbol{\theta})^H + \sigma\boldsymbol{I})). \tag{24}$$

We note that this limiting function depends on the emitter signals only through $\boldsymbol{S}$. It can be further shown that the unique minima of $V_{cml}(\boldsymbol{\theta})$ is when $\boldsymbol{\theta}$ is equal to its true value $\boldsymbol{\theta}_0$. Please refer to [5] for a simple proof of these facts. But then if the estimate of $\boldsymbol{\theta}$ is consistent, is it also asymptotically ($N \to \infty$) efficient ?

It turns out that the estimate of the DoA is not asymptotically ($N \to \infty$) efficient since the estimate of other parameters $(\boldsymbol{X}_N, \sigma)$ are not consistent. To see this fact note that asymptotically as $N \to \infty$, from (23) it follows that

$$\widehat{\boldsymbol{x}}(t)_{cml} = \boldsymbol{A}(\widehat{\boldsymbol{\theta}}_{cml})^{\dagger}\boldsymbol{y}(t) \overset{N\to\infty}{\longrightarrow} \boldsymbol{x}(t) + \boldsymbol{A}(\boldsymbol{\theta}_0)^{\dagger}\boldsymbol{w}(t) \neq \boldsymbol{x}(t). \tag{25}$$

Similarly, for $\widehat{\sigma}_{cml}$ we have

$$\widehat{\sigma}_{cml} \overset{N\to\infty}{\longrightarrow} \frac{\text{tr}(\boldsymbol{P}^{\perp}_{\boldsymbol{A}(\boldsymbol{\theta}_0)}\widehat{\boldsymbol{R}})}{m} = \frac{m-d}{m}\sigma \neq \sigma \text{ for finite } m. \tag{26}$$

Under the mild condition that

$$\boldsymbol{a}(\theta_i)^H\boldsymbol{a}(\theta_i) \overset{m\to\infty}{\longrightarrow} \infty \tag{27}$$

it can be shown that as $m \to \infty$ even the ML estimates of $\boldsymbol{X}_N$ and σ are consistent i.e., they are asymptotically ($N \to \infty$) equal to their true values.

For any unbiased estimate of the DoA $\widehat{\boldsymbol{\theta}}$, the CRLB on the error variance under CMA is given by

$$\mathbb{E}\Big((\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)^T\Big) \succcurlyeq \boldsymbol{B}_c \tag{28}$$

where

$$\left\{\boldsymbol{B}_c^{-1}\right\}_{i,j} = \frac{2N}{\sigma}\mathrm{Re}\left[\mathrm{tr}\left(\left\{\frac{\partial \boldsymbol{a}(\theta)}{\partial \theta}\Big|_{\theta=\boldsymbol{\theta}_{0_j}}\right\}^H \boldsymbol{P}^{\perp}_{\boldsymbol{A}(\boldsymbol{\theta}_0)}\left\{\frac{\partial \boldsymbol{a}(\theta)}{\partial \theta}\Big|_{\theta=\boldsymbol{\theta}_{0_i}}\right\}\boldsymbol{S}\right)\right] \tag{29}$$

Further, it can also be shown that asymptotically as $N \to \infty$, $\sqrt{N}(\widehat{\boldsymbol{\theta}}_{cml} - \boldsymbol{\theta}_0)$ is a zero mean Gaussian random vector with a covariance matrix given by

$$\boldsymbol{C}_{cml} = N\boldsymbol{B}_c + 2N^2\boldsymbol{B}_c\mathrm{Re}\left(\left(\boldsymbol{D}(\boldsymbol{\theta}_0)^H\boldsymbol{P}^{\perp}_{\boldsymbol{A}(\boldsymbol{\theta}_0)}\boldsymbol{D}(\boldsymbol{\theta}_0)\right)\odot\left(\boldsymbol{A}(\boldsymbol{\theta}_0)^H\boldsymbol{A}(\boldsymbol{\theta}_0)\right)^{-T}\right)\boldsymbol{B}_c. \tag{30}$$

From (30) it follows that for finite m asymptotically as $N \to \infty$ [5]

$$\frac{\boldsymbol{C}_{cml}}{N} \succ \boldsymbol{B}_c \text{ strict inequality} \tag{31}$$

However if $m \to \infty$ then under some mild conditions (27) it can be shown that $\frac{\boldsymbol{C}_{cml}}{N} = \boldsymbol{B}_c$.

Therefore we note that under CMA, the ML estimates of DoA are not asymptotically ($N \to \infty$) efficient, which is primarily due to the fact that the number of parameters to be estimated increases with N. The only way asymptotic efficiency can be achieved is by having a very large sensor array. This is in contrast to the ML estimate under UMA, which we had earlier shown to be asymptotically ($N \to \infty$) efficient even for finite m.

V. Asymptotic comparison between ML estimates under CMA and UMA

It can be shown that with emitter signals being second-order ergodic with covariance $\boldsymbol{S}$ the ML estimation error variance under UMA is asymptotically ($N \to \infty$) better than CMA. Note that, under UMA we assume a Gaussian distribution for the emitter signals.

$$\frac{\boldsymbol{C}_{cml}}{N} \succcurlyeq \frac{\boldsymbol{C}_{uml}}{N} = \boldsymbol{B}_u. \tag{32}$$

Please refer to [5] for proof.

VI. Are ML estimates of DoA under UMA always better than those under CMA ?

The relation in (32) assumed that under UMA the emitter signals are Gaussian distrbuted. Naturally an important question which arises here is whether the same relation is true irrespective of the distribution of the emitter signals.

It turns out that *indeed UMA results in better DoA estimates compared to CMA* for *arbitrary* second-order ergodic emitter signals. This is actually a consequence of the *asymptotic robustness* property of the ML estimates under both UMA and CMA, i.e., the asymptotic distribution of the parameter estimates is completely specified by $\lim_{N\to\infty}(1/N)\sum_{t=1}^{N}\boldsymbol{x}(t)\boldsymbol{x}(t)^H$. As shown in [5], [6], the actual emitter waveform sequence (or its distribution) is immaterial.

These results therefore provide strong justification for UMA being the more appropriate model for estimating DoA using sensor arrays.

REFERENCES

[1] B. Ottersten, M. Viberg, P. Stoica, and A. Nehorai, "Exact and Large Sample ML Techniques for Parameter Estimation and Detection in Array Processing," *in (S. Haykin et al., eds.), Radar Array Processing, New York, NY: Springer-Verlag, Inc.*, 1993, Chapter 4.

[2] J. F. Böhme, "Estimation of Spectral Parameters of Correlated Signals in Wavefileds," *Signal Processing*, 10:329–337, 1986.

[3] Y. H. Bresler, "Maximum Likelihood Estimation of Linearly Structured Covariance with Application to Antenna Array Processing," *In Proc. 4th ASSP Workshop on Spectrum Estimation and Modeling*, pp:172–175, Minneapolis, MN, Aug. 1988.

[4] A. G. Jaffer, "Maximum Likelihood Direction Finding of Stochastic Sources : A Separable Solution," *In proc. ICASSP 88*, vol. 5, pp: 2893–2896, New York, April 1988.

[5] P. Stoica and A. Nehorai, "Performance study of Conditional and Unconditional Direction-of-Arrival Estimation," *IEEE Trans. on Acoustics speech and Signal Processing*, vol. 38, no. 10, pp: 1783–1795, October 1990.

[6] B. Ottersten, M. Viberg, and T. Kailath, "Asymptotic Robustness of Sensor Array Processing Methods," *in Proc. ICCASP'90*, pp:2635–2638, Alburquerque, NM, April 1990.

[7] M. N. E. Korso, R. Boyer, A. Renaux and S. Marcos, "Conditional and Unconditional Cramer-Rao Bounds for Near-Field Source Localization," *IEEE Trans. on Signal Processing*, vol. 58, no. 5, pp:2901–2907, May 2010.

[8] J. D. Kalbfleisch, and D. A. Sprott, "Application of Likelihood Methods to Models Involving Large Numbers of Parameters," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 32, no. 2, pp:175–208, 1970.

[9] B. Hochwald, and A. Nehorai, "Concentrated Cramer-Rao Bound Expressions," *IEEE Trans. on Information Theory*, vol. 40, no. 2, pp:363–371, March 1994.