



HAL
open science

Performance bounds and statistical analysis of DOA estimation

Jean-Pierre Delmas

► **To cite this version:**

Jean-Pierre Delmas. Performance bounds and statistical analysis of DOA estimation. Array and statistical signal processing., 3, Elsevier, pp.719 - 764, 2015, Academic Press library in signal processing, 978-0-12-411597-2. 10.1016/B978-0-12-411597-2.00016-3 . hal-01285947v2

HAL Id: hal-01285947

<https://hal.science/hal-01285947v2>

Submitted on 25 May 2016

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.

Performance bounds and statistical analysis of DOA estimation

Jean Pierre Delmas

I. INTRODUCTION

Over the last three decades, many direction of arrival (DOA) estimation and source number detection methods have been proposed in the literature. Early studies on statistical performance were only based on extensive Monte Carlo experiments. Analytical performance evaluations, that allow one to evaluate the expected performance, as pioneered by [47], have since attracted much excellent research.

The earlier works were devoted to the statistical performance analysis of subspace-based algorithms. In particular the celebrated MUSIC algorithm has been extensively investigated (see e.g., [80], [82], [67], [103] among many others). But curiously, these works were based on first-order perturbations of the eigenvectors and eigenvalues of the sample covariance matrix, and thus involved very complicated derivations. Subsequently, [51] carried out a performance analysis of two eigenstructure-based DOA estimation algorithms, using a series expansion of the orthogonal projectors on the signal and noise subspaces, allowing considerable simplification of the previous approaches. Motivated by this point of view, several unified analyses of subspace-based algorithms have been presented (see e.g., [57], [40], [58]). In parallel to these works, a particular attention has been paid to the statistical performance of the exact and approximative maximum likelihood algorithms (ML), in relation to the celebrated Cramer-Rao bound (see e.g., [81], [63], and the tutorial [64] with the references therein).

The statistical performance analysis of the difficult and critical problem of the detection of the number of sources impinging on an array, has been based on principally standard techniques of the statistical detection literature. In particular, the information theoretical criteria and especially the minimum description length (MDL), as popularized in the signal processing literature by [99], have been analyzed (see e.g., [106] [33], [26]). Related to the DOA estimation accuracy and to the detection of the number of sources, the resolvability of closely spaced signals in terms of their parameters of interest have been also extensively studied (see e.g., [78], [59]).

The aim of this chapter is not to give a survey of all performance analysis of DOA estimation and source detection methods that have appeared in the literature, but rather, to provide a unified methodology introduced in [11] and then specialized to second-order in [19] to study the theoretical statistical performance of arbitrary DOA estimation and source number detection methods and to tackle the resolvability of closely spaced sources. To illustrate this framework, several examples are detailed such as the conventional MUSIC algorithm, the MDL criterion and the angular resolution limit based on the detection theory.

This chapter is organized as follows. Section II presents the mathematical model of the array output and introduce the basic assumptions. General statistical tools for performance bounds and statistical analysis of DOA estimation algorithms are given in Section III based on a functional approach providing a common unifying framework. Then, Section IV embarks on statistical performance analysis of beamforming-based, maximum likelihood and second-order algorithms with a particular attention paid to the subspace-based algorithms. In particular the robustness w.r.t. the Gaussian distribution, the independence and narrowband assumptions, and array modeling errors are considered. Finally some elements of statistical performance analysis of high-order algorithms complete this section. A glimpse into the detection of the number of sources is given in Section V where a performance analysis of the minimum description length (MDL) criterion is derived. Finally, Section VI is devoted to criteria for resolving two closely spaced sources.

Jean Pierre Delmas is with TELECOM SudParis, Département CITI, CNRS UMR 5157, 91011 Evry Cedex, France, e-mail:jean-pierre.delmas@it-sudparis.eu, phone: +(33).1.60.76.46.32, fax: +(33).1.60.76.44.33.

The following notations are used throughout this chapter: $o(\epsilon)$ and $O(\epsilon)$ denote quantities such that $\lim_{\epsilon \rightarrow 0} o(\epsilon)/\epsilon = 0$ and $|O(\epsilon)/\epsilon|$ is bounded in the neighborhood of $\epsilon = 0$, respectively.

II. MODELS AND BASIC ASSUMPTION

A. Parametric array model

Consider an array of M sensors arranged in an arbitrary geometry that receives the waveforms generated by P point sources (electromagnetic or acoustic). The output of each sensor is modeled as the response of a linear time-invariant bandpass system of bandwidth B . The impulse response of each sensor to a signal impinging on the array depends on the physical antenna structure, the receiver electronics and other antennas in the array through mutual coupling. The complex amplitudes $s_p(t)$ of these sources w.r.t. a carrier frequency f_0 are assumed to vary very slowly relative to the propagation time across the array (more precisely, the array aperture measured in wavelength, is much less than the inverse relative bandwidth f_0/B). This so-called narrowband assumption allows the time delays $\tau_{m,p}$ of the p th source at the m th sensor, relative to some fixed reference point, to be modeled as a simple phase-shift of the carrier frequency. If $\mathbf{n}(t)$ is the complex envelope of the additive noise, the complex envelope of the signals collected at the output of the sensors is given by applying the superposition principle for linear sensors by:

$$\mathbf{x}(t) = \sum_{p=1}^P \mathbf{a}(\boldsymbol{\theta}_p) s_p(t) + \mathbf{n}(t) = \mathbf{A}(\boldsymbol{\theta}) \mathbf{s}(t) + \mathbf{n}(t), \quad (1)$$

where $\mathbf{s}(t) \stackrel{\text{def}}{=} [s_1(t), \dots, s_P(t)]^T$ and $\boldsymbol{\theta}_p$ may include generally azimuth, elevation, range and polarization of the p th source. However, we will here assume that there is only one parameter per source, referred as the direction of arrival (DOA) θ . $\mathbf{a}(\theta_p)$ is the steering vector associated with the p th source. The array manifold, defined as the set $\{\mathbf{a}(\theta), \theta \in \Theta\}$ for some region Θ in DOA space, is perfectly known, either analytically or by measuring it in the field. It is further required for performance analysis that $\mathbf{a}(\theta)$ be continuously twice differentiable w.r.t. θ . $\mathbf{A}(\boldsymbol{\theta}) = [\mathbf{a}(\theta_1), \dots, \mathbf{a}(\theta_P)]$ is the $M \times P$ steering matrix with $\boldsymbol{\theta} = [\theta_1, \dots, \theta_P]^T$.

To illustrate the parameterization of the steering vector $\mathbf{a}(\theta)$, assume that the sources are in the far field of the array, and that the medium is non-dispersive, so that the waveforms can be approximated as planar. In this case, the m th component of $\mathbf{a}(\theta)$ is simply $g_m(\theta) e^{-i\mathbf{k}^T \mathbf{r}_m}$ where $g_m(\theta)$ is the directivity gain of the m th sensor, $\mathbf{k} \stackrel{\text{def}}{=} \frac{2\pi f_0}{c} \mathbf{u}$, c represents the speed of propagation, \mathbf{u} is a unit vector pointing in the direction of propagation and \mathbf{r}_m is the position of the m th sensor relative the origin of the different delays.

The by far most studied sensor geometry is that of uniform linear array (ULA), where the M sensors are assumed to be identical and omnidirectional over the DOA range of interest. Referenced w.r.t. the first sensor that is used as the origin, $g_m(\theta) = 1$ and $\mathbf{k}^T \mathbf{r}_m = (m-1) \frac{2\pi f_0}{c} d \sin(\theta) = (m-1) \frac{2\pi d}{\lambda_0} \sin(\theta)$, where λ_0 is the wavelength. To avoid any ambiguity, d must be less than or equal to $\frac{\lambda_0}{2}$. The standard ULA has $d = \frac{\lambda_0}{2}$ that ensures a maximum accuracy on the estimation of θ . In this case

$$\mathbf{a}(\theta) = [1, e^{i\pi \sin(\theta)}, \dots, e^{i(M-1)\pi \sin(\theta)}]^T. \quad (2)$$

B. Signal assumptions and problem formulation

Each vector observation $\mathbf{x}(t)$ is called a snapshot of the array output. Let the process $\mathbf{x}(t)$ be observed at N time instants $\{t_1, \dots, t_N\}$. $\mathbf{x}(t)$ is often sampled at a slow sampling frequency $1/T_s$ compared to the bandwidth of $\mathbf{x}(t)$ for which $\{\mathbf{x}(t)\}_{t_1, \dots, t_N}$ are independent. Temporal correlation between successive snapshots is generally not a problem, but implies that a larger number N of snapshots is needed for the same performance. We will prove in Subsection IV-C that the parameter that fixes the performance is not N , but the observation interval $T = NT_s$. The signals $\{s_p(t)\}_{p=1, \dots, P}$ and $\mathbf{n}(t)$ are assumed independent¹. For well calibrated arrays,

¹Note that only the uncorrelation assumption is required for second-order based algorithms, in contrast to fourth-order based algorithms, that require the independent assumption. However, this latter one simplifies the statistical performance analysis.

$\mathbf{n}(t)$ is often assumed to be dominated by thermal noise in the receivers, which can be well modeled as zero-mean temporally and spatially white circular Gaussian random process. In this case, $E[\mathbf{n}(t_i)\mathbf{n}^H(t_j)] = \sigma_n^2\delta_{i,j}\mathbf{I}$ and $E[\mathbf{n}(t_i)\mathbf{n}^T(t_j)] = \mathbf{0}$, for which the spatial covariance and spatial complementary covariance matrices are given by $\mathbf{R}_n \stackrel{\text{def}}{=} E[\mathbf{n}(t)\mathbf{n}^H(t)] = \sigma_n^2\mathbf{I}$ and $\mathbf{C}_n \stackrel{\text{def}}{=} E[\mathbf{n}(t)\mathbf{n}^T(t)] = \mathbf{0}$, respectively. A common, alternative model assumes that $\mathbf{n}(t)$ is spatially correlated where \mathbf{R}_n is known up to a scalar multiplicative term σ_n^2 , i.e., $\mathbf{R}_n = \sigma_n^2\mathbf{\Sigma}_n$ where $\mathbf{\Sigma}_n$ is a known definite positive matrix. In this case, $\mathbf{x}(t)$ can be pre-multiplied by an inverse square-root factor $\mathbf{\Sigma}_n^{-1/2}$ of $\mathbf{\Sigma}_n$, which renders the resulting noise spatially white and preserves model (1) by replacing the steering vectors $\mathbf{a}(\theta)$ by $\mathbf{\Sigma}_n^{-1/2}\mathbf{a}(\theta)$.

Two kind of assumptions are used for $\{s_p(t)\}_{p=1,\dots,P}$. In the first one, called stochastic or unconditional model (see e.g., [63] and [81]), $\{s_p(t)\}_{p=1,\dots,P}$ are assumed to be zero-mean random variables for which the most commonly used distribution is the circular Gaussian one with spatial covariance $\mathbf{R}_s \stackrel{\text{def}}{=} E[\mathbf{s}(t)\mathbf{s}^H(t)]$ and spatial complementary covariance $\mathbf{C}_s \stackrel{\text{def}}{=} E[\mathbf{s}(t)\mathbf{s}^T(t)] = \mathbf{0}$. \mathbf{R}_s is nonsingular for not fully correlated sources (called also noncoherent) or near-singular for highly correlated sources. In the case of coherent sources (specular multipath or smart jamming, where some signals impinging on the array of sensors can be sums of scaled and delayed versions of the others), \mathbf{R}_s is singular. In this chapter \mathbf{R}_s is usually assumed nonsingular. For these assumptions, the snapshots $\mathbf{x}(t)$ are zero-mean complex circular Gaussian distributed with covariance matrix

$$\mathbf{R}_x = \mathbf{A}(\boldsymbol{\theta})\mathbf{R}_s\mathbf{A}^H(\boldsymbol{\theta}) + \sigma_n^2\mathbf{I}. \quad (3)$$

This circular Gaussian assumption lies not only in the fact that circular Gaussian data are rather frequently encountered in applications, but also because optimal detection and estimation algorithms are much easier to deduce under this assumption. Furthermore, as will be discussed in Section IV, under rather general conditions and in large samples [86], the Gaussian CRB is the largest of all CRB matrices corresponding to different distributions of the sources of identical covariance matrix \mathbf{R}_s . This stochastic model can be extended by assuming that $\mathbf{s}(t)$ is arbitrarily distributed with finite fourth-order moments [19] including the case where $\mathbf{C}_s \neq \mathbf{0}$ associated with the second-order noncircular distributions.

A common alternative assumption, called deterministic or conditional model (see e.g., [63] and [81]) is used when the distribution of $\mathbf{s}(t)$ is unknown or/and clearly nonGaussian, for example in radar and radio communications. Here $\mathbf{s}(t)$ is nonrandom, i.e., the sequence $\{\mathbf{s}(t)\}_{t_1,\dots,t_N}$ is frozen in all realizations of the random snapshots $\{\mathbf{x}(t)\}_{t_1,\dots,t_N}$. Consequently, $\{\mathbf{s}(t)\}_{t_1,\dots,t_N}$ is considered as a complex unknown parameter in \mathbb{C}^{NP} . For this assumption, the snapshots $\mathbf{x}(t)$ are complex circular Gaussian distributed with mean $\mathbf{A}(\boldsymbol{\theta})\mathbf{s}(t)$ and covariance matrix $\sigma_n^2\mathbf{I}$.

With these preliminaries, the main DOA problem can now be formulated as follows: Given the observations, $\{\mathbf{x}(t)\}_{t_1,\dots,t_N}$ and the described model (1), detect the number P of incoming sources and estimate their DOAs $\{\theta_p\}_{p=1,\dots,P}$.

C. Parameter identifiability

Once the distribution of the observations $\{\mathbf{x}(t)\}_{t_1,\dots,t_N}$ has been fixed, the question of the identifiability of the parameters (including the DOA $\{\theta_p\}_{p=1,\dots,P}$) must be raised. For example, under the assumption of independent, zero-mean circular Gaussian distributed observations, all information in the measured data is contained in the covariance matrix \mathbf{R}_x (3). The question of parameter identifiability is thus reduced to investigating under which conditions \mathbf{R}_x determines the unknown parameters. Thus, if no a priori information on \mathbf{R}_s is available, the unknown parameter $\boldsymbol{\alpha}$ of \mathbf{R}_x contains the following $P + P^2 + 1$ real-valued parameters:

$$\boldsymbol{\alpha} = [\theta_1, \dots, \theta_P, [\mathbf{R}_s]_{1,1}, \dots, [\mathbf{R}_s]_{P,P}, \text{Re}([\mathbf{R}_s]_{2,1}), \text{Im}([\mathbf{R}_s]_{2,1}), \dots, \text{Re}([\mathbf{R}_s]_{P,P-1}), \text{Im}([\mathbf{R}_s]_{P,P-1}), \sigma_n^2]^T \quad (4)$$

and the parameter $\boldsymbol{\alpha}$ is identifiable if and only if $\mathbf{R}_x(\boldsymbol{\alpha}^{(1)}) = \mathbf{R}_x(\boldsymbol{\alpha}^{(2)}) \Rightarrow \boldsymbol{\alpha}^{(1)} = \boldsymbol{\alpha}^{(2)}$. To ensure this identifiability, it is necessary that $\mathbf{A}(\boldsymbol{\theta})$ be full column rank for any collection of P , distinct $\theta_p \in \Theta$. An array satisfying this assumption is said to be unambiguous. Notice that this requirement is problem-dependent and, therefore, has to be established for the specific array under study. For example, due to the Vandermonde structure of $\mathbf{a}(\theta)$ in the ULA case (2), it is straightforward to prove that the ULA is unambiguous if $\Theta = (-\pi/2, +\pi/2)$.

In the case where the rank of \mathbf{R}_s , that is the dimension of the linear space spanned by $\mathbf{s}(t)$ is known and equal to r , different conditions of identifiability has been given in the literature. In particular, the condition

$$P < \frac{M+r}{2} \quad (\text{which reduces to } P < M \text{ when } \mathbf{R}_s \text{ is nonsingular}) \quad (5)$$

has been proved to be sufficient [100] and practically necessary [61].

When $\mathbf{s}(t)$ are not circularly Gaussian distributed, the identifiability condition is generally much more involved. For example, when $\mathbf{s}(t)$ is noncircularly Gaussian distributed, $\mathbf{x}(t)$ is noncircularly Gaussian distributed as well with complementary covariance

$$\mathbf{C}_x = \mathbf{A}(\boldsymbol{\theta})\mathbf{C}_s\mathbf{A}^T(\boldsymbol{\theta}) \neq \mathbf{0} \quad (6)$$

and the distribution of the observations are now characterized by both \mathbf{R}_x and \mathbf{C}_x . Consequently, the condition of identifiability will be modified w.r.t. the circular case given in (5). This condition has not been presented in the literature, except for the particular case of uncorrelated and rectilinear (called also maximally improper) sources impinging on a ULA for which, the augmented covariance matrix $\mathbf{R}_{\tilde{\mathbf{x}}} \stackrel{\text{def}}{=} \mathbb{E}[\tilde{\mathbf{x}}(t)\tilde{\mathbf{x}}^H(t)]$ with $\tilde{\mathbf{x}}(t) \stackrel{\text{def}}{=} [\mathbf{x}^T(t), \mathbf{x}^H(t)]^T$ is given by

$$\mathbf{R}_{\tilde{\mathbf{x}}} = \sum_{p=1}^P \sigma_p^2 \mathbf{a}(\theta_p, \phi_p) \mathbf{a}^H(\theta_p, \phi_p) + \sigma_n^2 \mathbf{I}, \quad (7)$$

where $\mathbf{a}(\theta_p, \phi_p) \stackrel{\text{def}}{=} [\mathbf{a}^T(\theta_p), e^{-2i\phi_p} \mathbf{a}^H(\theta_p)]^T$ with ϕ_p is the second-order phase of noncircularity defined by

$$\mathbb{E}[s_p^2(t)] = e^{2i\phi_p} \mathbb{E}[s_p^2(t)] = e^{2i\phi_p} \sigma_p^2. \quad (8)$$

Due to the Vandermonde-like structure of the extended steering matrix $\mathbf{A}(\boldsymbol{\theta}, \boldsymbol{\phi}) \stackrel{\text{def}}{=} [\mathbf{a}(\theta_1, \phi_1), \dots, \mathbf{a}(\theta_P, \phi_P)]$, the condition of identifiability is now here $P < 2M - 1$.

Note that when $\mathbf{s}(t)$ is discrete distributed (for example when $s_p(t)$ are symbols $s_{p,k(p)}$ of a digital modulation taking q different values), the condition of identifiability is nontrivial despite the distribution of $\mathbf{x}(t)$ is a mixture of q^P circular Gaussian distributions of mean $\sum_{p=1}^P s_{p,k(p)} \mathbf{a}(\theta_p)$ and covariance $\sigma_n^2 \mathbf{I}$.

III. GENERAL STATISTICAL TOOLS FOR PERFORMANCE ANALYSIS OF DOA ESTIMATION

A. Performance analysis of a specific algorithm

1) *Functional analysis*: To study the statistical performance of any DOA's estimator (often called an algorithm as a succession of different steps), it is fruitful to adopt a functional analysis that consists in recognizing that the whole process of constructing the estimate $\hat{\boldsymbol{\theta}}_N$ is equivalent to defining a functional relation linking this estimate to the measurements from which it is inferred. As generally $\hat{\boldsymbol{\theta}}_N$ are functions of some statistics \mathbf{g}_N (assumed complex-valued vector in \mathbb{C}^L) deduced from $(\mathbf{x}(t))_{t_1, \dots, t_N}$, we have the following mapping

$$\{\mathbf{x}(t)\}_{t_1, \dots, t_N} \mapsto \mathbf{g}_N \xrightarrow{\text{alg}} \hat{\boldsymbol{\theta}}_N. \quad (9)$$

Many often, the statistics \mathbf{g}_N are sample moments or cumulants of $\mathbf{x}(t)$. The most common ones are second-order sample moments of $\mathbf{x}(t)$ deduced from the sample covariance and complementary covariance matrices $\mathbf{R}_{x,N} \stackrel{\text{def}}{=} \frac{1}{N} \sum_{n=1}^N \mathbf{x}(t_n) \mathbf{x}^H(t_n)$ and $\mathbf{C}_{x,N} \stackrel{\text{def}}{=} \frac{1}{N} \sum_{n=1}^N \mathbf{x}(t_n) \mathbf{x}^T(t_n)$, respectively. For nonGaussian symmetric sources distributions, even sample high-order cumulants of $\mathbf{x}(t)$ are also used, in particular the fourth-order sample cumulants deduced from the sample quadrivariance matrices $\mathbf{Q}_{x,N}$, $\mathbf{Q}'_{x,N}$ and $\mathbf{Q}''_{x,N}$ where $[\mathbf{Q}_x]_{i+(j-1)M, k+(l-1)M} \stackrel{\text{def}}{=} \text{Cum}(x_i(t), x_j^*(t), x_k^*(t), x_l(t))$, $[\mathbf{Q}'_x]_{i+(j-1)M, k+(l-1)M} \stackrel{\text{def}}{=} \text{Cum}(x_i(t), x_j^*(t), x_k(t), x_l(t))$ and $[\mathbf{Q}''_x]_{i+(j-1)M, k+(l-1)M} \stackrel{\text{def}}{=} \text{Cum}(x_i(t), x_j(t), x_k(t), x_l(t))$, estimated through the associated fourth and second-order sample moments. In these cases, the algorithms are called second-order, high-order and fourth-order algorithms, respectively.

The statistic \mathbf{g}_N generally satisfies two conditions:

(i) \mathbf{g}_N converges almost surely (from the strong law of large numbers) to $\mathbb{E}(\mathbf{g}_N)$ when N tends to infinity, that is a function of the DOAs and other parameters denoted $\mathbf{g}(\boldsymbol{\theta})$,

(ii) the DOAs $\boldsymbol{\theta}$ are identifiable from $\mathbf{g}(\boldsymbol{\theta})$, i.e., there exists a mapping $\mathbf{g}(\boldsymbol{\theta}) \mapsto \boldsymbol{\theta}$.

Furthermore, we assume that the algorithm \mathbf{alg} satisfies $\mathbf{alg}[\mathbf{g}(\boldsymbol{\theta})] = \boldsymbol{\theta}$ for all $\boldsymbol{\theta} \in \Theta$. Consequently the functional dependence $\hat{\boldsymbol{\theta}}_N = \mathbf{alg}(\mathbf{g}_N)$ constitutes a particular extension of the mapping $\mathbf{g}(\boldsymbol{\theta}) \mapsto \boldsymbol{\theta}$ in the neighborhood of $\mathbf{g}(\boldsymbol{\theta})$ that characterizes all algorithm based on the statistic \mathbf{g}_N .

Note that for circular Gaussian stochastic and deterministic models of the sources, the likelihood functions of the measurements depend on $\{\mathbf{x}(t)\}_{t_1, \dots, t_N}$ through only the sample covariance $\mathbf{R}_{x,N}$, and therefore the algorithms called respectively stochastic maximum likelihood (SML) and deterministic maximum likelihood (DML) algorithms are second-order algorithms [64]. The SML algorithm has been extended to noncircular Gaussian sources, for which the ML algorithm is built from both $\mathbf{R}_{x,N}$ and $\mathbf{C}_{x,N}$ [21].

However, due to their complexity, many suboptimal algorithms with much lower computational requirements have been proposed in the literature. Among them, many algorithms are based on the noise (or signal) orthogonal projector $\mathbf{\Pi}_{x,N}$ onto the noise (or signal) subspace associated with the sample covariance $\mathbf{R}_{x,N}$. These algorithms are called subspace-based algorithms. The most celebrated is the MUSIC algorithm that offers a good trade-off between performance and computational costs. Its statistical performance has been thoroughly studied in the literature (see e.g., [47], [80], [34], [31]). In these cases, the mapping (9) becomes

$$\{\mathbf{x}(t)\}_{t_1, \dots, t_N} \mapsto \mathbf{R}_{x,N} \mapsto \mathbf{\Pi}_{x,N} \xrightarrow{\mathbf{alg}} \hat{\boldsymbol{\theta}}_N, \quad (10)$$

where the mapping \mathbf{alg} characterizes the specific subspace-based algorithm. Some of these algorithms have been extended for noncircular sources through subspace-based algorithms based on $(\mathbf{\Pi}_{x,N}, \mathbf{\Pi}'_{x,N})$ or $\mathbf{\Pi}_{\tilde{x},N}$ where $\mathbf{\Pi}'_{x,N}$ and $\mathbf{\Pi}_{\tilde{x},N}$ are the orthogonal projectors onto the noise subspace associated with the sample complementary covariance $\mathbf{C}_{x,N}$ and the sample augmented covariance $\mathbf{R}_{\tilde{x},N} \stackrel{\text{def}}{=} \frac{1}{N} \sum_{n=1}^N \tilde{\mathbf{x}}(t_n) \tilde{\mathbf{x}}^H(t_n)$ with $\tilde{\mathbf{x}}(t_n) \stackrel{\text{def}}{=} (\mathbf{x}^T(t_n), \mathbf{x}^H(t_n))^T$, respectively [3].

2) *Asymptotic distribution of statistics:* Due to the nonlinearity of model (1) w.r.t. the DOA's parameter, the performance analysis of detectors for the number of sources and the DOA's estimation procedures are not possible for a finite number N of snapshots. But in many cases, asymptotic performance analyses are available when the number N of measurements, the signal-to-noise ratio (SNR) (see e.g., [71]) or the number of sensors M converges to infinity (see e.g., [92]). In practice N , SNR and M are naturally finite and thus available results in the asymptotic regime are approximations, whose domain of validity are specified through Monte Carlo simulations. We will consider in this chapter, only asymptotic properties w.r.t. N and thus, the presented results will be only valid in practice when $N \gg M$. When N is of the same order of magnitude than M , although very large, the approximations given by the asymptotic regime w.r.t. N are generally very bad.

To derive the asymptotic distribution, covariance and bias of estimated DOAs w.r.t. the number N of measurements, we first need to specify the asymptotic distribution of some statistics \mathbf{g}_N .

For the second-order statistics

$$\mathbf{g}_N = \text{vec}(\mathbf{R}_{x,N}, \mathbf{C}_{x,N}) = \frac{1}{N} \sum_{n=1}^N \begin{bmatrix} \mathbf{x}^*(t_n) \otimes \mathbf{x}(t_n) \\ \mathbf{x}(t_n) \otimes \mathbf{x}(t_n) \end{bmatrix},$$

where $\text{vec}(\cdot)$ and \otimes denote, respectively, the vectorization operator that turns a matrix into a vector by stacking the columns of the matrix one below another and the standard Kronecker product of matrices, closed-form expressions of the covariance $\mathbb{E}[(\mathbf{g}_N - \mathbf{g})(\mathbf{g}_N - \mathbf{g})^H]$ and complementary covariance $\mathbb{E}[(\mathbf{g}_N - \mathbf{g})(\mathbf{g}_N - \mathbf{g})^T]$ matrices (where $\mathbf{g} \stackrel{\text{def}}{=} \mathbf{g}(\boldsymbol{\theta})$ for short), and their asymptotic distributions² have been given [25] for independent

²Throughout this chapter $\mathcal{N}_R(\mathbf{m}; \mathbf{R})$, $\mathcal{N}_C(\mathbf{m}; \mathbf{R})$ and $\mathcal{N}_C(\mathbf{m}; \mathbf{R}, \mathbf{C})$ denote the real, circular complex, arbitrary complex Gaussian distribution, respectively, with mean \mathbf{m} , covariance \mathbf{R} and complementary covariance \mathbf{C} .

measurements, fourth-order arbitrary distributed sources and Gaussian distributed noise:

$$\begin{aligned} \mathbb{E} [(\mathbf{g}_N - \mathbf{g})(\mathbf{g}_N - \mathbf{g})^H] &= \frac{1}{N} \begin{pmatrix} \mathbf{R}_{R_x} & \mathbf{R}_{R_x, C_x} \\ \mathbf{R}_{R_x, C_x}^H & \mathbf{R}_{C_x} \end{pmatrix} \\ \mathbb{E} [(\mathbf{g}_N - \mathbf{g})(\mathbf{g}_N - \mathbf{g})^T] &= \frac{1}{N} \begin{pmatrix} \mathbf{C}_{R_x} & \mathbf{C}_{R_x, C_x} \\ \mathbf{C}_{R_x, C_x}^T & \mathbf{C}_{C_x} \end{pmatrix}, \\ \sqrt{N} (\text{vec}(\mathbf{R}_{x,N}, \mathbf{C}_{x,N}) - \text{vec}(\mathbf{R}_x, \mathbf{C}_x)) &\xrightarrow{\mathcal{L}} \mathcal{N}_C \left(\mathbf{0}; \begin{pmatrix} \mathbf{R}_{R_x} & \mathbf{R}_{R_x, C_x} \\ \mathbf{R}_{R_x, C_x}^H & \mathbf{R}_{C_x} \end{pmatrix}, \begin{pmatrix} \mathbf{C}_{R_x} & \mathbf{C}_{R_x, C_x} \\ \mathbf{C}_{R_x, C_x}^T & \mathbf{C}_{C_x} \end{pmatrix} \right), \end{aligned} \quad (11)$$

with

$$\begin{aligned} \mathbf{R}_{R_x} &= \mathbf{R}_x^* \otimes \mathbf{R}_x + \mathbf{K}(\mathbf{C}_x \otimes \mathbf{C}_x^*) + (\mathbf{A}^* \otimes \mathbf{A}) \mathbf{Q}_s (\mathbf{A}^T \otimes \mathbf{A}^H) \\ \mathbf{R}_{C_x} &= \mathbf{R}_x \otimes \mathbf{R}_x + \mathbf{K}(\mathbf{R}_x \otimes \mathbf{R}_x) + (\mathbf{A} \otimes \mathbf{A}) \mathbf{Q}_s''' (\mathbf{A}^H \otimes \mathbf{A}^H) \\ \mathbf{C}_{R_x} &= \mathbf{R}_{R_x} \mathbf{K} \\ \mathbf{C}_{C_x} &= \mathbf{C}_x \otimes \mathbf{C}_x + \mathbf{K}(\mathbf{C}_x \otimes \mathbf{C}_x) + (\mathbf{A} \otimes \mathbf{A}) \mathbf{Q}_s'' (\mathbf{A}^T \otimes \mathbf{A}^T) \\ \mathbf{R}_{R_x, C_x} &= \mathbf{C}_x^* \otimes \mathbf{R}_x + \mathbf{K}(\mathbf{R}_x \otimes \mathbf{C}_x^*) + (\mathbf{A}^* \otimes \mathbf{A}) \mathbf{Q}_s'''' (\mathbf{A}^H \otimes \mathbf{A}^H) \\ \mathbf{C}_{R_x, C_x} &= \mathbf{R}_x^* \otimes \mathbf{C}_x + \mathbf{K}(\mathbf{C}_x \otimes \mathbf{R}_x^*) + (\mathbf{A}^* \otimes \mathbf{A}) \mathbf{Q}_s' (\mathbf{A}^T \otimes \mathbf{A}^T), \end{aligned} \quad (12)$$

where $\mathbf{A} \stackrel{\text{def}}{=} \mathbf{A}(\boldsymbol{\theta})$ for short and \mathbf{K} denotes the vec-permutation matrix which transforms $\text{vec}(\mathbf{C})$ to $\text{vec}(\mathbf{C}^T)$ for any square matrix \mathbf{C} . \mathbf{Q}_s , \mathbf{Q}_s' and \mathbf{Q}_s'' are defined as for $\mathbf{x}(t)$ defined previously and $[\mathbf{Q}_s''']_{i+(j-1)P, k+(l-1)P} \stackrel{\text{def}}{=} \text{Cum}(s_i(t), s_j(t), s_k^*(t), s_l^*(t))$, $[\mathbf{Q}_s''''']_{i+(j-1)P, k+(l-1)P} \stackrel{\text{def}}{=} \text{Cum}(s_i(t), s_j^*(t), s_k^*(t), s_l^*(t))$. Note that the asymptotic distribution of $\mathbf{R}_{x,N}$ has been extended to non independent measurements with arbitrary distributed sources and noise of finite fourth-order moments with \mathbf{R}_n arbitrarily structured in [18] in [19].

Consider now the noise orthogonal projector $\mathbf{g}_N = \text{vec}(\boldsymbol{\Pi}_{x,N})$. Its asymptotic distribution is deduced from the standard first-order perturbation for orthogonal projectors [46] (see also [51]):

$$\delta(\boldsymbol{\Pi}_{x,N}) = -\boldsymbol{\Pi}_x \delta(\mathbf{R}_{x,N}) \mathbf{S}^\# - \mathbf{S}^\# \delta(\mathbf{R}_{x,N}) \boldsymbol{\Pi}_x + o(\delta(\mathbf{R}_{x,N})), \quad (13)$$

where $\delta(\boldsymbol{\Pi}_{x,N}) \stackrel{\text{def}}{=} \boldsymbol{\Pi}_{x,N} - \boldsymbol{\Pi}_x$, $\delta(\mathbf{R}_{x,N}) \stackrel{\text{def}}{=} \mathbf{R}_{x,N} - \mathbf{R}_x$ and $\mathbf{S}^\#$ is the Moore-Penrose inverse of $\mathbf{S} = \mathbf{A}(\boldsymbol{\theta}) \mathbf{R}_s \mathbf{A}^H(\boldsymbol{\theta})$. The remainder in (13) is a standard $o(\delta(\mathbf{R}_{x,N}))$ for a realization of the random matrix $\mathbf{R}_{x,N}$, but an $o_p(\delta(\mathbf{R}_{x,N}))$ if $\mathbf{R}_{x,N}$ is considered as random. The relation (13) proves that \mathbf{g}_N is differentiable w.r.t. $\text{vec}(\mathbf{R}_{x,N})$ in the neighborhood of $\text{vec}(\mathbf{R}_x)$ and its differential matrix (called also Jacobian matrix) evaluated at $\text{vec}(\mathbf{R}_x)$ is

$$\mathbf{D}_{R_x, \boldsymbol{\Pi}_x} = - \left(\mathbf{S}^{*\#} \otimes \boldsymbol{\Pi}_x + \boldsymbol{\Pi}_x^* \otimes \mathbf{S}^\# \right). \quad (14)$$

Then using the standard theorem of continuity (see e.g., [75, th.B, p.124]) on regular functions of asymptotically Gaussian statistics, the asymptotic behaviors of $\boldsymbol{\Pi}_{x,N}$ and $\mathbf{R}_{x,N}$ are directly related:

$$\sqrt{N} (\text{vec}(\boldsymbol{\Pi}_{x,N}) - \text{vec}(\boldsymbol{\Pi}_x)) \xrightarrow{\mathcal{L}} \mathcal{N}_C(\mathbf{0}; \mathbf{R}_{\boldsymbol{\Pi}_x}, \mathbf{R}_{\boldsymbol{\Pi}_x} \mathbf{K}), \quad (15)$$

where $\mathbf{R}_{\boldsymbol{\Pi}_x}$ is given for independent measurements, fourth-order arbitrary distributed sources and Gaussian distributed noise, using (12) by:

$$\mathbf{R}_{\boldsymbol{\Pi}_x} = \mathbf{D}_{R_x, \boldsymbol{\Pi}_x} \mathbf{R}_{R_x} \mathbf{D}_{R_x, \boldsymbol{\Pi}_x}^H = \boldsymbol{\Pi}_x^* \otimes \mathbf{U} + \mathbf{U}^* \otimes \boldsymbol{\Pi}_x, \quad (16)$$

with $\mathbf{U} = \sigma_n^2 \mathbf{S}^\# \mathbf{R}_s \mathbf{S}^\#$. We see that $\mathbf{R}_{\boldsymbol{\Pi}_x}$ does not depend on \mathbf{C}_s and the quadrivariances of the sources. Consequently, all subspace-based algorithms are robust to the distribution and to the noncircularity of the sources; i.e., the asymptotic performances are those of the standard complex circular Gaussian case. Note that the asymptotic distribution of $(\boldsymbol{\Pi}_{x,N}, \boldsymbol{\Pi}'_{x,N})$ and $\boldsymbol{\Pi}_{\bar{x},N}$ have also been derived under the same assumptions in [3], where it is proved that they do not depend on the quadrivariances of the sources, as well. The asymptotic distributions of $\boldsymbol{\Pi}_{x,N}$, $(\boldsymbol{\Pi}_{x,N}, \boldsymbol{\Pi}'_{x,N})$ and $\boldsymbol{\Pi}_{\bar{x},N}$ will allow us to derive the statistical performance of arbitrary

subspace-based algorithms based on these orthogonal projectors in the Section IV-D.

Note that the second-order expansion of $\mathbf{\Pi}_{x,N}$ w.r.t. $\mathbf{R}_{x,N}$ has been used in [51] to analyse the behavior of the root-MUSIC and root-min-norm algorithms dedicated to ULA, but is useless as far as we are concerned by the asymptotic distribution of the DOAs alone, as it has been specified in [3], where an extension of the root-MUSIC algorithm to noncircular sources has been proposed.

Finally, consider now the asymptotic distribution of the signal eigenvalues of $\mathbf{R}_{x,N}$ that is useful for the statistical performance analysis of information theoretic criteria (whose MDL criterion popularized by Wax and Kailath [99] is one of the most successful), for the detection of the number P of sources. Let $\lambda_1, \dots, \lambda_P, \lambda_{P+1} = \sigma_n^2, \dots, \lambda_M = \sigma_n^2$ denote the eigenvalues of \mathbf{R}_x , ordered in decreasing order and $\mathbf{v}_1, \dots, \mathbf{v}_P$ the associated eigenvectors (defined up to a multiplicative unit modulus complex number) of the signal subspace. Then, suppose that for a "small enough" perturbation $\mathbf{R}_{x,N} - \mathbf{R}_x$, the largest P associated eigenvalues of the sample covariance $\mathbf{R}_{x,N}$ are $\lambda_{1,N} > \dots > \lambda_{P,N}$. It is proved in [26], extending the work by [47] to arbitrary distributed independent measurements (1) with finite fourth-order moment, not necessarily circular and Gaussian, the following convergence in distribution.

$$\sqrt{N} (\boldsymbol{\lambda}_N - \boldsymbol{\lambda}) \xrightarrow{\mathcal{L}} \mathcal{N}_R(\mathbf{0}; \mathbf{R}_\lambda), \quad (17)$$

with $\boldsymbol{\lambda}_N = [\lambda_{1,N}, \dots, \lambda_{P,N}]^T$, $\boldsymbol{\lambda} = [\lambda_1, \dots, \lambda_P]^T$ and $[\mathbf{R}_\lambda]_{i,j} = \lambda_i^2 \delta_{i,j} + |\lambda_{i,j}|^2 + \lambda_{i,i,j,j}$ for $i, j = 1, \dots, P$, $\delta_{i,j}$ is the Kronecker delta, $\lambda_{i,j} \stackrel{\text{def}}{=} \mathbf{v}_i^H \mathbf{C}_x \mathbf{v}_j^*$ and $\lambda_{i,j,k,l} \stackrel{\text{def}}{=} (\mathbf{v}_i^T \otimes \mathbf{v}_j^H) \mathbf{Q}_x (\mathbf{v}_k^* \otimes \mathbf{v}_l)$. In contrast to the circular Gaussian distribution [47], we see that the estimated eigenvalues $\{\lambda_{i,N}\}_{i=1,\dots,P}$ are no longer asymptotically mutually independent. Furthermore, it is proved in [26] that for $i, j = 1, \dots, P$:

$$\mathbb{E}[\lambda_{i,N}] = \lambda_i + \frac{1}{N} \sum_{1 \leq k \neq i \leq M} \frac{\lambda_i \lambda_k + |\lambda_{i,k}|^2 + \lambda_{i,k,i,k}}{\lambda_i - \lambda_k} + o\left(\frac{1}{N}\right) \quad (18)$$

$$\text{Cov}[\lambda_{i,N}, \lambda_{j,N}] = \frac{1}{N} (\lambda_i^2 \delta_{i,j} + |\lambda_{i,j}|^2 + \lambda_{i,i,j,j}) + o\left(\frac{1}{N}\right). \quad (19)$$

We note that these results are also valid for the augmented covariance matrix $\mathbf{R}_{\tilde{x},N}$ where M and P are replaced by $2M$ and the rank of $\mathbf{R}_{\tilde{x},N} - \sigma_n^2 \mathbf{I}$, respectively.

3) *Asymptotic distribution of estimated DOA*: In the following, we consider arbitrary DOA algorithms that are in practice "regular" enough³. More specifically, we assume that the mapping \mathbf{alg} is \mathbb{R} -differentiable w.r.t. $\mathbf{g}_N \in \mathbb{C}^L$ in the neighborhood of $\mathbf{g}(\boldsymbol{\theta})$, i.e.,

$$\widehat{\boldsymbol{\theta}}_N = \mathbf{alg}(\mathbf{g}_N) = \mathbf{alg}(\mathbf{g}) + \mathbf{D}_{g,\boldsymbol{\theta}}^{\text{alg}}(\mathbf{g}_N - \mathbf{g}) + \mathbf{D}_{g,\boldsymbol{\theta}}^{\text{alg}*}(\mathbf{g}_N - \mathbf{g})^* + o\|\mathbf{g}_N - \mathbf{g}\|, \quad (20)$$

with $\mathbf{alg}(\mathbf{g}) = \boldsymbol{\theta}$ and $P \times L$ matrix $\mathbf{D}_{g,\boldsymbol{\theta}}^{\text{alg}}$ is the \mathbb{R} -differential matrix (Jacobian) of the mapping $\mathbf{g}_N \xrightarrow{\text{alg}} \widehat{\boldsymbol{\theta}}_N$ evaluated at $\mathbf{g}(\boldsymbol{\theta})$. In practice, this matrix is derived from the chain rule by decomposing the algorithm as successive simpler mappings, and in each of these mapping, this matrix is simply deduced from first-order expansions. Then, applying a simple extension of the standard theorem of continuity [75, th.B, p.124] (also called Δ -method), it is straightforwardly proved the following convergence in distribution:

$$\sqrt{N} (\widehat{\boldsymbol{\theta}}_N - \boldsymbol{\theta}) \xrightarrow{\mathcal{L}} \mathcal{N}_R(\mathbf{0}; \mathbf{R}_\theta) \quad \text{with} \quad \mathbf{R}_\theta = 2 \left[\mathbf{D}_{g,\boldsymbol{\theta}}^{\text{alg}} \mathbf{R}_g (\mathbf{D}_{g,\boldsymbol{\theta}}^{\text{alg}})^H + \text{Re} \left(\mathbf{D}_{g,\boldsymbol{\theta}}^{\text{alg}} \mathbf{C}_g (\mathbf{D}_{g,\boldsymbol{\theta}}^{\text{alg}})^T \right) \right], \quad (21)$$

where \mathbf{R}_g and \mathbf{C}_g are the covariance and the complementary covariance matrices of the asymptotic distribution of the statistics \mathbf{g}_N . We note that for subspace-based algorithms and second-order algorithms based on $\mathbf{R}_{x,N}$ or $\mathbf{R}_{\tilde{x},N}$, $\mathbf{g}_N^* = \mathbf{K} \mathbf{g}_N$ (because the orthogonal projector matrices and the covariance matrices are Hermitian structured), and generally for statistics \mathbf{g}_N that contain all conjugate of its components, the mapping \mathbf{alg} is \mathbb{C} -differentiable w.r.t. \mathbf{g}_N in the neighborhood of $\mathbf{g}(\boldsymbol{\theta})$ and (20) and (21) become respectively:

$$\widehat{\boldsymbol{\theta}}_N = \mathbf{alg}(\mathbf{g}_N) = \mathbf{alg}(\mathbf{g}) + \mathbf{D}_{g,\boldsymbol{\theta}}^{\text{alg}}(\mathbf{g}_N - \mathbf{g}) + o\|\mathbf{g}_N - \mathbf{g}\|, \quad (22)$$

³This is the case, for example when $\widehat{\boldsymbol{\theta}}_N$ maximizes w.r.t. $\boldsymbol{\alpha}$, a real-valued function $f(\boldsymbol{\alpha}, \mathbf{g}_N)$ that is twice- \mathbb{R} differentiable w.r.t. $\boldsymbol{\alpha}$ and \mathbf{g}_N .

where now, $\mathbf{D}_{g,\theta}^{\text{alg}}$ is the \mathbb{C} -differential matrix of the mapping $\mathbf{g}_N \xrightarrow{\text{alg}} \widehat{\boldsymbol{\theta}}_N$ evaluated at $\mathbf{g}(\boldsymbol{\theta})$ and

$$\sqrt{N} (\widehat{\boldsymbol{\theta}}_N - \boldsymbol{\theta}) \xrightarrow{\mathcal{L}} \mathcal{N}_R(\mathbf{0}; \mathbf{R}_\theta) \quad \text{with} \quad \mathbf{R}_\theta = \mathbf{D}_{g,\theta}^{\text{alg}} \mathbf{R}_g (\mathbf{D}_{g,\theta}^{\text{alg}})^H. \quad (23)$$

4) *Asymptotic covariance and bias*: Under additional regularities of the algorithm alg , that are generally satisfied, the covariance of $\widehat{\boldsymbol{\theta}}_N$ is given by

$$\text{Cov}(\widehat{\boldsymbol{\theta}}_N) = \frac{1}{N} \mathbf{R}_\theta + o\left(\frac{1}{N}\right). \quad (24)$$

Using a second-order expansion of $\text{alg}(\mathbf{g}_N)$ and $\mathbb{C}\mathbb{R}$ -calculus, where alg is assumed to be twice- \mathbb{R} -differentiable, the bias is given by

$$\mathbb{E}(\widehat{\boldsymbol{\theta}}_N) - \boldsymbol{\theta} = \frac{1}{2N} \begin{bmatrix} \text{Tr}(\mathbf{R}_{\tilde{g}} \mathbf{H}_{\tilde{g},\theta,1}^{\text{alg}}) \\ \vdots \\ \text{Tr}(\mathbf{R}_{\tilde{g}} \mathbf{H}_{\tilde{g},\theta,P}^{\text{alg}}) \end{bmatrix} + o\left(\frac{1}{N}\right), \quad (25)$$

where $\mathbf{H}_{\tilde{g},\theta,k}^{\text{alg}} = \frac{\partial}{\partial \tilde{\mathbf{g}}} \left(\frac{\partial \text{alg}}{\partial \tilde{\mathbf{g}}} \right)^H = \begin{bmatrix} \mathbf{H}_{g,\theta,k}^{(1)} & \mathbf{H}_{g,\theta,k}^{(2)} \\ \mathbf{H}_{g,\theta,k}^{(2)*} & \mathbf{H}_{g,\theta,k}^{(1)*} \end{bmatrix}$ is the complex augmented Hessian matrix [74, A2.3]

of the k th component of the function alg at point $\mathbf{g}(\boldsymbol{\theta})$ and $\mathbf{R}_{\tilde{g}} = \begin{bmatrix} \mathbf{R}_g & \mathbf{C}_g \\ \mathbf{C}_g^* & \mathbf{R}_g^* \end{bmatrix}$ is the augmented covariance of the asymptotic distribution of \mathbf{g}_N . In the particular case where alg is twice- \mathbb{C} -differentiable (see e.g., the examples given for \mathbb{C} -differentiable algorithms (22)), i.e.,

$$\widehat{\boldsymbol{\theta}}_N = \text{alg}(\mathbf{g}_N) = \text{alg}(\mathbf{g}) + \mathbf{D}_{g,\theta}^{\text{alg}} (\mathbf{g}_N - \mathbf{g}) + \frac{1}{2} [\mathbf{I}_P \otimes (\mathbf{g}_N - \mathbf{g})^H] \begin{bmatrix} \mathbf{H}_{g,\theta,1}^{\text{alg}} \\ \vdots \\ \mathbf{H}_{g,\theta,P}^{\text{alg}} \end{bmatrix} [\mathbf{g}_N - \mathbf{g}] + o\|\mathbf{g}_N - \mathbf{g}\|^2, \quad (26)$$

(25) reduces to

$$\mathbb{E}(\widehat{\boldsymbol{\theta}}_N) - \boldsymbol{\theta} = \frac{1}{2N} \begin{bmatrix} \text{Tr}(\mathbf{R}_g \mathbf{H}_{g,\theta,1}^{\text{alg}}) \\ \vdots \\ \text{Tr}(\mathbf{R}_g \mathbf{H}_{g,\theta,P}^{\text{alg}}) \end{bmatrix} + o\left(\frac{1}{N}\right). \quad (27)$$

We note that relations (24), (25) and (27) are implicitly used in the signal processing literature by simple first and second-order expansions of the estimate $\widehat{\boldsymbol{\theta}}_N$ w.r.t. the involved statistics without checking any necessary mathematical conditions concerning the remainder terms of the first and second-order expansions. In fact these conditions are very difficult to prove for the involved mappings $\mathbf{g}_N \xrightarrow{\text{alg}} \widehat{\boldsymbol{\theta}}_N$. For example, the following necessary conditions are given in [56, Th. 4.2.2] for second-order algorithms: (i) the measurements $\{\mathbf{x}(t)\}_{t_1, \dots, t_N}$ are independent with finite eighth moments, (ii) the mapping $\mathbf{g}_N \xrightarrow{\text{alg}} \widehat{\boldsymbol{\theta}}_N$ is four times \mathbb{R} -differentiable, (iii) the fourth derivative of this mapping and those of its square are bounded. These assumptions that do not depend on the distribution of the measurements are very strong, but fortunately (24), (25) and (27) continue to hold in many cases in which these assumptions are not satisfied, in particular for Gaussian distributed data (see, e.g., [56, Ex. 4.2.2]).

In practice, (24), (25) and (27) show that the mean square error (MSE)

$$\mathbb{E}\|\widehat{\boldsymbol{\theta}}_N - \boldsymbol{\theta}\|^2 = \|\mathbb{E}(\widehat{\boldsymbol{\theta}}_N) - \boldsymbol{\theta}\|^2 + \text{Tr}[\text{Cov}(\widehat{\boldsymbol{\theta}}_N)] \quad (28)$$

is then also of order $1/N$. Its main contribution comes from the variance term, since the square of the bias is of order $1/N^2$. But as empirically observed, this bias contribution may be significant when SNR or N is not sufficiently large. However, there are very few contributions in the literature, that have derived closed-form bias expressions. Among them, [104] has considered the bias of the MUSIC algorithm, whose derivation ought to be simplified by using the asymptotic distribution of the orthogonal projector $\mathbf{\Pi}_{x,N}$, rather than those of the sample signal eigenvectors $(\mathbf{e}_{1,N}, \dots, \mathbf{e}_{P,N})$.

B. Cramer-Rao bounds (CRB)

The accuracy measures of performance in terms of covariance and bias of any algorithm, described in the previous section may be of limited interest, unless one has an idea of what the best possible performance is. An important measure of how well a particular DOA finding algorithm performs is the mean square error (MSE) matrix $\mathbb{E}[(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T]$ of the estimation error $\hat{\boldsymbol{\theta}}_N - \boldsymbol{\theta}$. Among the lower bounds on this matrix, the celebrated Cramer-Rao bound (CRB) is by far the most commonly used. We note that this CRB is indeed deduced from the CRB on the complete unknown parameter $\boldsymbol{\alpha}$ of the parametrized DOA model, for example, given by (4) for the circular Gaussian stochastic model. Furthermore, rigorously speaking, this CRB ought to be only used for unbiased estimators and under sufficiently regular distributions of the measurements. Fortunately, these technical conditions are satisfied in practice and due to the property that the bias contribution is often weak w.r.t. the variance term in the mean square error (28) for $N \gg 1$, the CRB that lower bounds the covariance matrix of any unbiased estimators is used to lower bound the MSE matrix of any asymptotically unbiased estimator⁴

$$\mathbb{E}[(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha})(\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha})^T] \geq \text{CRB}(\boldsymbol{\alpha}) \quad (29)$$

with $\text{CRB}(\boldsymbol{\alpha})$ is given under weak regularity conditions by:

$$\text{CRB}(\boldsymbol{\alpha}) = \mathbf{FIM}^{-1}(\boldsymbol{\alpha}), \quad (30)$$

where $\mathbf{FIM}(\boldsymbol{\alpha})$ is the Fisher information matrix (FIM) given elementwise by

$$[\mathbf{FIM}(\boldsymbol{\alpha})]_{k,l} = -\mathbb{E} \left[\left(\frac{\partial^2 \log p(\mathbf{x}; \boldsymbol{\alpha})}{\partial \alpha_k \partial \alpha_l} \right) \right] \quad (31)$$

associated with the probability density function $p(\mathbf{x}; \boldsymbol{\alpha})$ of the measurements $\mathbf{x} = [\mathbf{x}^T(t_1), \dots, \mathbf{x}^T(t_N)]^T$.

The main reason for the interest of this CRB is that it is often asymptotically (when the amount N of data is large) tight, i.e., there exist algorithms, such that the stochastic maximum likelihood (ML) estimator (see IV-B), whose covariance matrices asymptotically achieve this bound. Such estimators are said to be asymptotically efficient. However, at low SNR and/or at low number N of snapshots, the CRB is not achieved and is overly optimistic. This is due to the fact that estimators are generally biased in such non-asymptotic cases. For these reasons, other lower bounds are available in the literature, that are more relevant to lower bound the MSE matrices. But unfortunately, their closed-form expressions are much more complex to derive and are generally non interpretable (see e.g., the Weiss-Weinstein bound in [97]).

In practice, closed-form expressions of the FIM (31) are difficult to obtain for arbitrary distributions of the sources and noise. In general, the involved integrations of (31) are solved numerically by replacing the expectations by arithmetical averages over a large number of computer generated measurements. But for Gaussian distributions, there are a plethora of closed-form expressions of $\text{CRB}(\boldsymbol{\theta})$ in the literature. And the reason of the popularity of this CRB is the simplicity of the FIM for Gaussian distributions of \mathbf{x} .

1) *Gaussian stochastic case:* One way to derive closed-form expressions of $\text{CRB}(\boldsymbol{\theta})$ is to use the extended Slepian-Bangs [77], [7] formula, where the FIM (31) is given elementwise by

$$[\mathbf{FIM}(\boldsymbol{\alpha})]_{k,l} = 2\text{Re} \left[\left(\frac{\partial \mathbf{m}_x}{\partial \alpha_k} \right)^H \mathbf{R}_x^{-1} \frac{\partial \mathbf{m}_x}{\partial \alpha_l} \right] + \text{Tr} \left[\frac{\partial \mathbf{R}_x}{\partial \alpha_k} \mathbf{R}_x^{-1} \frac{\partial \mathbf{R}_x}{\partial \alpha_l} \mathbf{R}_x^{-1} \right] \quad (32)$$

for a circular⁵ Gaussian $\mathcal{N}_C(\mathbf{m}_x; \mathbf{R}_x)$ distribution of \mathbf{x} . But there are generally difficulties to derive compact matrix expressions of the CRB for DOA parameters alone given by

$$\text{CRB}(\boldsymbol{\theta}) = [\mathbf{FIM}^{-1}(\boldsymbol{\alpha})]_{(1:P,1:P)}$$

⁴Note that for finite N , the estimator $\hat{\boldsymbol{\alpha}}$ is always biased and (29) does not apply. Additionally, biased estimators may exist whose MSE matrices are smaller than the CRB (see, e.g., [83]).

⁵Note that this Slepian-Bangs formula has been extended to noncircular Gaussian $\mathcal{N}_C(\mathbf{m}_x; \mathbf{R}_x, \mathbf{C}_x)$ distribution in [22] where (32) becomes $[\mathbf{FIM}(\boldsymbol{\alpha})]_{k,l} = \left(\frac{\partial \mathbf{m}_{\tilde{x}}}{\partial \alpha_k} \right)^H \mathbf{R}_{\tilde{x}}^{-1} \frac{\partial \mathbf{m}_{\tilde{x}}}{\partial \alpha_l} + \frac{1}{2} \text{Tr} \left[\frac{\partial \mathbf{R}_{\tilde{x}}}{\partial \alpha_k} \mathbf{R}_{\tilde{x}}^{-1} \frac{\partial \mathbf{R}_{\tilde{x}}}{\partial \alpha_l} \mathbf{R}_{\tilde{x}}^{-1} \right]$ with $\mathbf{m}_{\tilde{x}} \stackrel{\text{def}}{=} (\mathbf{m}_x^T, \mathbf{m}_x^H)^T$ and $\mathbf{R}_{\tilde{x}} \stackrel{\text{def}}{=} \begin{bmatrix} \mathbf{R}_x & \mathbf{C}_x \\ \mathbf{C}_x^* & \mathbf{R}_x^* \end{bmatrix}$.

with $\boldsymbol{\alpha} = (\boldsymbol{\theta}^T, \boldsymbol{\beta}^T)^T$ where $\boldsymbol{\beta}$ gathers all the nuisance parameters (in many applications, only the DOAs are of interest). Another way, based on the asymptotic efficiency of the ML estimator (under certain regularity conditions) has been used to indirectly derive the CRB on the DOA parameter alone (see IV-B).

For the circular Gaussian stochastic model of the sources introduced in Subsection II-B, compact matrix expressions of $\text{CRB}(\boldsymbol{\theta})$ have been given in the literature, when no a priori information is available on the structure of the spatial covariance \mathbf{R}_s of the sources. For example, Stoica *et al.* [87] have derived the following expression for one parameter per source and uniform white noise (i.e., $\mathbf{R}_n = \sigma_n^2 \mathbf{I}$)

$$\text{CRB}_{\text{CG}}(\boldsymbol{\theta}) = \frac{\sigma_n^2}{2N} \left\{ \text{Re} \left[(\mathbf{D}^H \boldsymbol{\Pi}_x \mathbf{D}) \odot (\mathbf{R}_s \mathbf{A}^H \mathbf{R}_x^{-1} \mathbf{A} \mathbf{R}_s)^T \right] \right\}^{-1}, \quad (33)$$

where \odot denotes the Hadamard product (i.e., element-wise multiplication), $\boldsymbol{\Pi}_x$ is the orthogonal projector on the noise subspace, i.e., $\boldsymbol{\Pi}_x = \boldsymbol{\Pi}_A^\perp \stackrel{\text{def}}{=} \mathbf{I} - \mathbf{A}(\mathbf{A}^H \mathbf{A})^{-1} \mathbf{A}^H$ and $\mathbf{D} \stackrel{\text{def}}{=} \left[\frac{d\mathbf{a}(\theta_1)}{d\theta_1}, \dots, \frac{d\mathbf{a}(\theta_P)}{d\theta_P} \right]$. We note the surprising fact that when the sources are known to be coherent (i.e., \mathbf{R}_s singular), the associated Gaussian CRB $\text{CRB}_{\text{CG}}(\boldsymbol{\theta})$ that includes this prior, keeps the same expression (33) [85].

As is well known, the importance of this Gaussian CRB formula lies in the fact that circular Gaussian data are rather frequently encountered in applications. Another important point is that under rather general conditions that will be specified in Subsection IV-B, the circular complex Gaussian CRB matrix (33) is the largest of all CRB matrices among the class of arbitrary complex distributions of the sources with given covariance matrix \mathbf{R}_x (see, e.g., [86, p. 293]). Note that many extensions of (33) have been given. For example this formula has been extended to several parameters per source (see e.g., [4, Appendix.D]), to nonuniform white noise (i.e., $\mathbf{R}_n = \text{Diag}[\sigma_1^2, \dots, \sigma_M^2]$) and unknown parameterized noise field (i.e., $\mathbf{R}_n = \boldsymbol{\Sigma}(\boldsymbol{\sigma})$) in [65] and [105], [39], respectively. Due to the domination of the Gaussian distribution, these bounds have often been denoted in the literature as stochastic CRB (e.g., in [63]) or unconditional CRB (e.g., in [81]), without specifying the involved distribution.

Furthermore, all these closed-form expressions of the CRB have been extended to the noncircular Gaussian stochastic model of the sources in [22], [2] and [4, Appendix.D], given associated $\text{CRB}_{\text{NCG}}(\boldsymbol{\theta})$ expressions satisfying

$$\text{CRB}_{\text{NCG}}(\boldsymbol{\theta}) \leq \text{CRB}_{\text{CG}}(\boldsymbol{\theta})$$

corresponding to the same covariance matrix \mathbf{R}_s . For example, for a single source, with one parameter θ_1 , $\text{CRB}_{\text{NCG}}(\theta_1)$ decreases monotonically as the second-order noncircularity rate γ_1 (defined by $\text{E}|s_1^2(t)| = \gamma_1 e^{2i\phi_1} \text{E}[s_1^2(t)]$) and satisfying $0 \leq \gamma_1 \leq 1$) increases from 0 to 1, for which we have, respectively,

$$\text{CRB}_{\text{CG}}(\theta_1) = \frac{1}{N} \left(\frac{1}{h_1} \left[\frac{\sigma_n^2}{\sigma_1^2} + \frac{1}{\|\mathbf{a}(\theta_1)\|^2} \frac{\sigma_n^4}{\sigma_1^4} \right] \right), \quad \text{CRB}_{\text{NCG}}(\theta_1) = \frac{1}{N} \left(\frac{1}{h_1} \left[\frac{\sigma_n^2}{\sigma_1^2} + \frac{1}{2\|\mathbf{a}(\theta_1)\|^2} \frac{\sigma_n^4}{\sigma_1^4} \right] \right), \quad (34)$$

where h_1 is the purely geometrical factor $2 \frac{d\mathbf{a}^H(\theta_1)}{d\theta_1} \boldsymbol{\Pi}_{\mathbf{a}_1}^\perp \frac{d\mathbf{a}(\theta_1)}{d\theta_1}$ with $\boldsymbol{\Pi}_{\mathbf{a}_1}^\perp \stackrel{\text{def}}{=} \mathbf{I}_M - \frac{\mathbf{a}(\theta_1) \mathbf{a}^H(\theta_1)}{\|\mathbf{a}(\theta_1)\|^2}$.

If the source covariance \mathbf{R}_s is constrained to have a specific structure, (i.e., if a prior on \mathbf{R}_s is taken into account), a specific expression of $\text{CRB}_{\text{CG}}(\boldsymbol{\theta})$, which integrates this prior ought to be derived, to assess the performance of an algorithm that uses this prior. But unfortunately, the derivation of $\text{CRB}_{\text{CG}}(\boldsymbol{\theta})$ is very involved and lacks any engineering insight. For example, when it is known that the sources are uncorrelated, the expression given in [44, theorem 1] of $\text{CRB}_{\text{CG}}(\boldsymbol{\theta})$ includes a matrix \mathbf{B} , defined as any matrix, whose columns span the null space of $[\mathbf{a}^*(\theta_1) \otimes \mathbf{a}(\theta_1), \dots, \mathbf{a}^*(\theta_P) \otimes \mathbf{a}(\theta_P)]^H$. And to the best of our knowledge no closed-form expression of $\text{CRB}_{\text{CG}}(\boldsymbol{\theta})$ has been published in the important case of coherent sources, when the rank of \mathbf{R}_s is fixed strictly smaller than P .

Finally, note that the scalar field modeling one component of electromagnetic field or acoustic pressure (1) has been extended to vector fields with vector sensors, where associated stochastic CRBs for the DOA (azimuth and elevation) alone have been derived and analyzed for a single source. In particular, the electromagnetic (six electric and magnetic field components) and acoustic (three velocity components and pressure) fields have been considered in [62] and [42], respectively.

2) *Gaussian deterministic case*: For the deterministic model of the sources introduced in Subsection II-B, the unknown parameter α of \mathbf{R}_x is now

$$\alpha = \left[\theta_1, \dots, \theta_P, \{ \text{Re}[\mathbf{s}^T(t_n)], \text{Im}[\mathbf{s}^T(t_n)] \}_{n=1, \dots, N}, \sigma_n^2 \right]^T. \quad (35)$$

Applying the extended Slepian-Bangs formula (32) to the circular Gaussian $\mathcal{N}_C \left(\begin{bmatrix} \mathbf{A}\mathbf{s}(t_1) \\ \vdots \\ \mathbf{A}\mathbf{s}(t_N) \end{bmatrix}; \sigma_n^2 \mathbf{I}_{NM} \right)$ distribution of \mathbf{x} , Stoica *et al.* [81] have obtained the following CRB for the DOA alone: $\text{CRB}_{\text{Det}}(\boldsymbol{\theta}) = \frac{\sigma_n^2}{2N} \{ \text{Re} [\mathbf{D}^H \boldsymbol{\Pi}_x \mathbf{D} \odot \mathbf{R}_{s,N}] \}^{-1}$, where $\mathbf{R}_{s,N} \stackrel{\text{def}}{=} \frac{1}{N} \sum_{n=1}^N \mathbf{s}(t_n) \mathbf{s}^H(t_n)$. Furthermore, it was proved in [80] that $\text{CRB}_{\text{Det}}(\boldsymbol{\theta})$ decreases monotonically with increasing N (and M). This implies, that if the sources $\mathbf{s}(t_n)$ are second-order ergodic sequences, $\mathbf{R}_{s,N}$ has a limit \mathbf{R}_s when N tends to infinity, and we obtain for large N , the following expression denoted in the literature as deterministic CRB or conditional CRB (e.g., in [81])

$$\text{CRB}_{\text{Det}}(\boldsymbol{\theta}) \approx \frac{\sigma_n^2}{2N} \{ \text{Re} [(\mathbf{D}^H \boldsymbol{\Pi}_x \mathbf{D}) \odot \mathbf{R}_s] \}^{-1}. \quad (36)$$

Finally, we remark that the CRB for near-field DOA localization has been much less studied than the far-field one. To the best of our knowledge, only papers [28], [8] and [27] have given and analyzed closed-form expressions of the stochastic and deterministic CRB, and furthermore in the particular case of a single source for specific arrays. For a ULA where the DOA parameters are the azimuth θ and the range r , based on the DOA algorithms, the steering vector (2) has been approximated in [28] by

$$[\mathbf{a}(\theta, r)]_{m=1, \dots, M} = e^{i(\omega(m-1) + \phi(m-1)^2)},$$

where ω and ϕ are the so-called electric angles connected to the physical parameters θ and r by $\omega = 2\pi \frac{d}{\lambda_0} \sin(\theta)$ and $\phi = \pi \frac{d^2}{\lambda_0 r} \cos^2(\theta)$. Then in [8], the exact propagation model

$$[\mathbf{a}(\theta, r)]_{m=1, \dots, M} = e^{i \frac{2\pi r}{\lambda_0} \left(\sqrt{1 + \frac{2(m-1)d \sin(\theta)}{r} + \frac{(m-1)^2 d^2}{r^2}} - 1 \right)},$$

has been used, that has revealed interesting features and interpretations not shown in [28]. Very recently, the uniform circular array (UCA) has been investigated in [27] in which the exact propagation model is now:

$$[\mathbf{a}(\theta, \phi, r)]_{m=1, \dots, M} = e^{i \frac{2\pi r}{\lambda_0} \left(1 - \sqrt{1 - 2 \frac{r_0}{r} \cos(\theta - \frac{2\pi(m-1)}{M}) \sin(\phi) + \frac{r_0^2}{r^2}} \right)},$$

where r_0 , θ and ϕ denote the radius of the UCA, the azimuth and the elevation of the source. Note that in contrast to the closed-form expressions given in [28] and [8], the ones given in [27] relate the near and far-field CRB on the azimuth and elevation by very simple expressions.

3) *Non Gaussian case*: The stochastic CRB for the DOA appears to be prohibitive to compute for non-Gaussian sources. To cope with this difficulty, the deterministic model for the sources has been proposed for its simplicity. But in contrast to the stochastic ML estimator, the corresponding deterministic (or conditional) ML method does not asymptotically achieve this deterministic CRB, because the deterministic likelihood function does not meet the required regularity conditions (see Subsection IV-B). Consequently, this deterministic CRB is only a nonattainable lower bound on the covariance of any unbiased DOA estimator for arbitrary nonGaussian distributions of the sources. So, it is useful to have explicit expressions of the stochastic CRB under non-Gaussian distributions.

To the best of our knowledge, such stochastic CRBs have only been given in the case of binary phase-shift keying (BPSK), quaternary phase-shift keying (QPSK) signal waveforms [23] and then, to arbitrary L -ary square QAM constellation [9], and for a single source only. In these works, it is assumed Nyquist shaping and ideal sample timing apply so that the intersymbol interference at each symbol spaced sampling instance can be ignored. In the absence of frequency offset but with possible phase offset, the signals at the output of the matched filter can be represented as $s_1(t) = \sigma_1^2 e^{i\phi_1} \epsilon_1(t)$, where $\{\epsilon_1(t)\}_{t_1, \dots, t_N}$ are independent identically

distributed random symbols taking values ± 1 for BPSK symbols and $\{\pm(2k-1)a \pm i(2l-1)a\}_{l,k=1,\dots,2^{q-1}}$ with $L = 2^{2q}$ for L -ary square QAM symbols, where $2a$ is the intersymbol distance in the I/Q plane, which is adjusted such that $E|\epsilon_1(t)|^2 = 1$. For these discrete sources, the unknown parameter of this stochastic model is

$$\boldsymbol{\alpha} = [\theta_1, \phi_1, \sigma_1^2, \sigma_n^2]^T$$

and it has been proved in [23] and [9] that the parameters (θ_1, ϕ_1) and (σ_1^2, σ_n^2) are decoupled in the associated FIM. This allows one to derive closed-form expressions of the so called non-data-aided (NDA) CRBs on the parameter θ_1 alone. In particular, it has been proved [23] that for a BPSK and QPSK source, that is respectively rectilinear and second-order circular, we have

$$\frac{\text{CRB}_{\text{BPSK}}(\theta_1)}{\text{CRB}_{\text{NCG}}(\theta_1)} = \frac{1}{(1-g(\rho))(1+\frac{1}{2\rho})} \quad \text{and} \quad \frac{\text{CRB}_{\text{QPSK}}(\theta_1)}{\text{CRB}_{\text{CG}}(\theta_1)} = \frac{1}{(1-g(\frac{\rho}{2}))(1+\frac{1}{\rho})}, \quad (37)$$

where $\text{CRB}_{\text{NCG}}(\theta_1)$ and $\text{CRB}_{\text{CG}}(\theta_1)$ are given by (34) and with $\rho \stackrel{\text{def}}{=} \frac{M\sigma_1^2}{\sigma_n^2}$ and g is the following decreasing function of ρ : $g(\rho) \stackrel{\text{def}}{=} \frac{e^{-\rho}}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \frac{e^{-\frac{u^2}{2}}}{\cosh(u\sqrt{2\rho})} du$. (37) is illustrated in Fig.1 for a ULA of M sensors spaced a half-wavelength apart. We see from this figure that the CRBs under the non-circular [resp. circular] complex Gaussian distribution are tight upper bounds on the CRBs under the BPSK [resp. QPSK] distribution at very low and very high SNRs only. Finally, note that among the numerous results of [23] and [9], these stochastic NDA CRBs have been compared with those obtained with different a priori knowledge. In particular, it has been proved that in the presence of any unknown phase offset (i.e., non-coherent estimation), the ultimate achievable performance on the NDA DOA estimates holds almost the same irrespectively of the modulation order L . However, the NDA CRBs obtained in the absence of phase offset (i.e., coherent estimation) vary, in the high SNR region, from one modulation order to another.

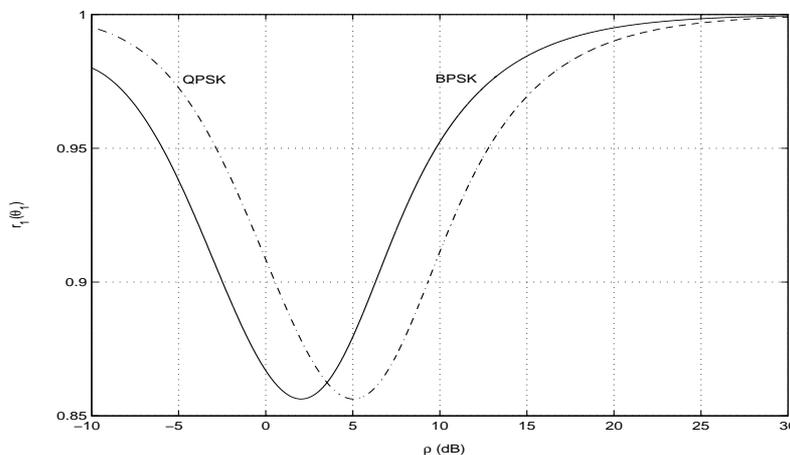


Fig.1 Ratios $r_1(\theta_1) \stackrel{\text{def}}{=} \frac{\text{CRB}_{\text{BPSK}}(\theta_1)}{\text{CRB}_{\text{NCG}}(\theta_1)}$ and $r_1(\theta_1) \stackrel{\text{def}}{=} \frac{\text{CRB}_{\text{QPSK}}(\theta_1)}{\text{CRB}_{\text{CG}}(\theta_1)}$ as a function of $\rho \stackrel{\text{def}}{=} \frac{M\sigma_1^2}{\sigma_n^2}$.

Finally note that the ML estimation of the DOAs of these discrete sources has been proposed [52], where the maximization of the ML criterion (which is rather involved) is iteratively carried out by the expectation maximization (EM) algorithm. Adapted to the distribution of these sources, this approach allows one to account for any arbitrary noise covariance \mathbf{R}_n as soon as $\mathbf{n}(t)$ is Gaussian distributed.

C. Asymptotically minimum variance bounds (AMVB)

To assess the performance of an algorithm based on a specific statistic \mathbf{g}_N built on $\{\mathbf{x}(t)\}_{t_1, \dots, t_N}$, it is interesting to compare the asymptotic covariance \mathbf{R}_θ (21) or (23) to an attainable lower bound that depends on the statistic \mathbf{g}_N only. The asymptotically minimum variance bound (AMVB) is such a bound. Furthermore, we note that the CRB appears to be prohibitive to compute for non-Gaussian sources and noise, except in simple cases and consequently this AMVB can be used as a useful benchmark against which potential estimates $\hat{\boldsymbol{\theta}}_N$

are tested. To extend the derivations of Porat and Friedlander [68] concerning this AMVB to complex-valued measurements, two additional conditions to those introduced in Subsection III-A1 must be satisfied:

(iii) the involved function \mathbf{alg} that defines the considered algorithm must be \mathbb{C} -differentiable, i.e., must satisfy (22). In practice, it is sufficient to add conjugate components to all complex-valued components of \mathbf{g} , as in example (41);

(iv) the covariance \mathbf{R}_g of the asymptotic distribution of \mathbf{g}_N must be nonsingular. To satisfy this latter condition, the components of \mathbf{g}_N that are random variables, must be asymptotically linearly independent. Consequently the redundancies in \mathbf{g}_N must be withdrawn.

Under these four conditions, the covariance matrix \mathbf{R}_θ of the asymptotic distribution of any estimator $\hat{\boldsymbol{\theta}}_N$ built on the statistics \mathbf{g}_N is bounded below by $(\mathbf{G}^H(\boldsymbol{\theta})\mathbf{R}_g^{-1}\mathbf{G}(\boldsymbol{\theta}))^{-1}$:

$$\mathbf{R}_\theta = \mathbf{D}_{g,\theta}^{\mathbf{alg}}\mathbf{R}_g(\mathbf{D}_{g,\theta}^{\mathbf{alg}})^H \geq (\mathbf{G}^H(\boldsymbol{\theta})\mathbf{R}_g^{-1}\mathbf{G}(\boldsymbol{\theta}))^{-1}, \quad (38)$$

where $\mathbf{G}(\boldsymbol{\theta})$ is the $L \times P$ matrix $\frac{d\mathbf{g}(\boldsymbol{\theta})}{d\boldsymbol{\theta}}$.

Furthermore, this lowest bound $\text{AMVB}_{\mathbf{g}_N}(\boldsymbol{\theta}) \stackrel{\text{def}}{=} (\mathbf{G}^H(\boldsymbol{\theta})\mathbf{R}_g^{-1}\mathbf{G}(\boldsymbol{\theta}))^{-1}$ is asymptotically tight, i.e., there exists an algorithm \mathbf{alg} whose covariance of its asymptotic distribution satisfies (38) with equality. The following nonlinear least square algorithm is an AMV second-order algorithm.

$$\hat{\boldsymbol{\theta}}_N = \arg \min_{\boldsymbol{\alpha} \in \Theta^P} [\mathbf{g}_N - \mathbf{g}(\boldsymbol{\alpha})]^H \mathbf{R}_g^{-1}(\boldsymbol{\alpha}) [\mathbf{g}_N - \mathbf{g}(\boldsymbol{\alpha})], \quad (39)$$

where we have emphasized here the dependence of \mathbf{R}_g on the unknown DOA $\boldsymbol{\alpha}$. In practice, it is difficult to optimize the nonlinear function (39), where it involves the computation of $\mathbf{R}_g^{-1}(\boldsymbol{\alpha})$. Porat and Friedlander proved for the real case in [35] that the lowest bound (38) is also obtained if an arbitrary weakly consistent estimate $\mathbf{R}_{g,N}$ of $\mathbf{R}_g(\boldsymbol{\alpha})$ is used in (39), giving the simplest algorithm:

$$\hat{\boldsymbol{\theta}}_N = \arg \min_{\boldsymbol{\alpha} \in \Theta^P} [\mathbf{g}_N - \mathbf{g}(\boldsymbol{\alpha})]^H \mathbf{R}_{g,N} [\mathbf{g}_N - \mathbf{g}(\boldsymbol{\alpha})]. \quad (40)$$

This property has been extended to the complex case in [69].

This AMVB and AMV algorithm have been applied to second-order algorithms that exploit both $\mathbf{R}_{x,N}$ and $\mathbf{C}_{x,N}$ in [21]. In this case, to fulfill the previously mentioned conditions (i-iv), the second-order statistics \mathbf{g}_N are given by

$$\mathbf{g}_N = \begin{bmatrix} \text{vec}(\mathbf{R}_{x,N}) \\ \mathbf{v}(\mathbf{C}_{x,N}) \\ \mathbf{v}(\mathbf{C}_{x,N}^*) \end{bmatrix}, \quad (41)$$

where $\mathbf{v}(\cdot)$ denotes the operator obtained from $\text{vec}(\cdot)$ by eliminating all supradiagonal elements of a matrix. Finally, note that these AMVB and AMV DOA finding algorithm have been also derived for fourth-order statistics by splitting the measurements and statistics \mathbf{g}_N into its real and imaginary parts in [69].

D. Relations between AMVB and CRB: projector statistics

The AMVB based on any statistics is generally lower bounded by the CRB because this later bound concerns arbitrary functions of the measurements $\{\mathbf{x}(t)\}_{t_1, \dots, t_N}$. But it has been proved in [4], that the AMVB associated with the different estimated projectors $\boldsymbol{\Pi}_{x,N}$, $(\boldsymbol{\Pi}_{x,N}, \boldsymbol{\Pi}'_{x,N})$ and $\boldsymbol{\Pi}_{\tilde{x},N}$ introduced in Subsection III-A2, which are functions of the second-order statistics of the measurements, attains the stochastic CRB in the case of circular or noncircular Gaussian signals. Consequently, there always exist asymptotically efficient subspace-based DOA algorithms in the Gaussian context.

To prove this asymptotic efficiency, i.e.,

$$\text{AMVB}_{\text{vec}(\boldsymbol{\Pi}_{x,N})}(\boldsymbol{\theta}) = \text{CRB}_{\text{CG}}(\boldsymbol{\theta}) \quad (42)$$

and

$$\text{AMVB}_{\text{vec}(\boldsymbol{\Pi}_{x,N}, \boldsymbol{\Pi}'_{x,N})}(\boldsymbol{\theta}) = \text{AMVB}_{\text{vec}(\boldsymbol{\Pi}_{\tilde{x},N})}(\boldsymbol{\theta}) = \text{CRB}_{\text{NCG}}(\boldsymbol{\theta}), \quad (43)$$

the condition (iv) of Subsection III-C that is not satisfied [1] for these statistics ought to be extended and consequently the results (38) and (39) must be modified as well, because here \mathbf{R}_g is singular.

In this singular case, it has been proved [1] that if the condition (iv) in the necessary conditions (i-iv) is replaced by the new condition $\text{Span}(\mathbf{G}(\boldsymbol{\theta})) \subset \text{Span}(\mathbf{R}_g(\boldsymbol{\theta}))$, (38) and (39) becomes respectively

$$\mathbf{R}_\theta = \mathbf{D}_{g,\theta}^{\text{alg}} \mathbf{R}_g (\mathbf{D}_{g,\theta}^{\text{alg}})^H \geq \left(\mathbf{G}^H(\boldsymbol{\theta}) \mathbf{R}_g^\# \mathbf{G}(\boldsymbol{\theta}) \right)^{-1} \quad (44)$$

and

$$\hat{\boldsymbol{\theta}}_N = \arg \min_{\boldsymbol{\alpha} \in \Theta^P} [\mathbf{g}_N - \mathbf{g}(\boldsymbol{\alpha})]^H \mathbf{R}_g^\#(\boldsymbol{\alpha}) [\mathbf{g}_N - \mathbf{g}(\boldsymbol{\alpha})]. \quad (45)$$

And it is proved that the three statistics $\text{vec}(\boldsymbol{\Pi}_{x,N})$, $\text{vec}(\boldsymbol{\Pi}_{x,N}, \boldsymbol{\Pi}'_{x,N})$ and $\text{vec}(\boldsymbol{\Pi}_{\bar{x},N})$ satisfy the conditions (i,ii,iii,v) and thus satisfy results (44) and (45).

Finally, note that this efficiency property of the orthogonal projectors extends to the model of spatially correlated noise, for which $\mathbf{R}_n = \sigma_n^2 \boldsymbol{\Sigma}_n$ where $\boldsymbol{\Sigma}_n$ is a known positive definite matrix. In this case, for example, the orthogonal projector $\boldsymbol{\Pi}_{x_w,N}$ defined after whitening

$$\{\mathbf{x}(t)\}_{t_1, \dots, t_N} \mapsto \{\mathbf{x}_w(t)\}_{t_1, \dots, t_N} \stackrel{\text{def}}{=} \{\boldsymbol{\Sigma}_n^{-1/2} \mathbf{x}(t)\}_{t_1, \dots, t_N} \mapsto \mathbf{R}_{x_w,N} = \frac{1}{N} \sum_{n=1}^N \mathbf{x}_w(t_n) \mathbf{x}_w^H(t_n) \mapsto \boldsymbol{\Pi}_{x_w,N}$$

satisfies

$$\text{AMVB}_{\text{vec}(\boldsymbol{\Pi}_{x_w,N})}(\boldsymbol{\theta}) = \text{CRB}_{\text{CG}}^w(\boldsymbol{\theta}) = \frac{\sigma_n^2}{2N} \left\{ \text{Re} \left[(\mathbf{D}^H \boldsymbol{\Pi}_{x_w} \mathbf{D}) \odot (\mathbf{R}_s \mathbf{A}^H \mathbf{R}_x^{-1} \mathbf{A} \mathbf{R}_s)^T \right] \right\}^{-1},$$

where $\boldsymbol{\Pi}_{x_w} \stackrel{\text{def}}{=} \boldsymbol{\Sigma}_n^{-1} - \boldsymbol{\Sigma}_n^{-1} \mathbf{A} (\mathbf{A}^H \boldsymbol{\Sigma}_n^{-1} \mathbf{A})^{-1} \boldsymbol{\Sigma}_n^{-1} \mathbf{A}^H$ is insensitive to the choice of the square root $\boldsymbol{\Sigma}_n^{1/2}$ of $\boldsymbol{\Sigma}_n$, and is no longer a projection matrix.

IV. ASYMPTOTIC DISTRIBUTION OF ESTIMATED DOA

We are now specifying in this section the asymptotic statistical performances of the main DOA algorithms that may be classified into three main categories, namely beamforming-based, maximum likelihood and moments-based algorithms.

A. Beamforming-based algorithms

Among the so-called beamforming-based algorithms, also referred to as low-resolution, compared to the parametric algorithms, the conventional (Bartlett) beamforming and Capon beamforming are the most referenced representatives of this family. These algorithms do not make any assumption on the covariance structure of the data, but the functional form of the steering vector $\mathbf{a}(\theta)$ is assumed perfectly known. These estimators $\hat{\boldsymbol{\theta}}_N$ are given by the P highest (supposed isolated) maximizer and minimizer in α of the respective following criteria

$$\mathbf{a}^H(\alpha) \hat{\mathbf{R}}_x \mathbf{a}(\alpha) \quad \text{and} \quad \mathbf{a}^H(\alpha) \widehat{\mathbf{R}}_x^{-1} \mathbf{a}(\alpha) \quad (46)$$

where $\hat{\mathbf{R}}_x$ is the unbiased sample estimate $\mathbf{R}_{x,N}$ and $\widehat{\mathbf{R}}_x^{-1}$ is either the biased estimate $\mathbf{R}_{x,N}^{-1}$ or the unbiased estimate $[(N-M)/N] \mathbf{R}_{x,N}^{-1}$ (that both give the same estimate $\hat{\boldsymbol{\theta}}_N$). Note that these algorithms extend to d parameters per source, where α is replaced by $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_d)$ in (46).

For arbitrary noise field, (i.e., arbitrary noise covariance \mathbf{R}_n) and/or an arbitrary number P of sources, the estimate $\hat{\boldsymbol{\theta}}_N$ given by these two algorithms are nonconsistent, i.e.,

$$\lim_{N \rightarrow \infty} \hat{\boldsymbol{\theta}}_N \neq \boldsymbol{\theta}$$

and asymptotically biased. The asymptotic bias $\text{AsBias}(\boldsymbol{\theta})$ can be straightforwardly derived by a second-order expansion of the criterion $\mathbf{a}^H(\alpha) \mathbf{R}_x^\epsilon \mathbf{a}(\alpha)$ around each true values $(\theta_p)_{p=1, \dots, P}$ (with $\epsilon = +1$ [resp., $\epsilon = -1$] for the conventional [resp. Capon] algorithm), but noting that $\lim_{N \rightarrow \infty} \mathbb{E}(\hat{\theta}_{p,N})$ is a maximizer or minimizer

$\bar{\theta}_p$ of $\mathbf{a}^H(\alpha)\mathbf{R}_x\mathbf{a}(\alpha)$ or $\mathbf{a}^H(\alpha)\mathbf{R}_x^{-1}\mathbf{a}(\alpha)$, respectively. The following value is obtained [91]

$$\text{AsBias}(\theta_p) \stackrel{\text{def}}{=} \lim_{N \rightarrow \infty} \text{E}(\hat{\theta}_{p,N}) - \theta_p = -\frac{\text{Re}[\mathbf{a}'^H(\theta_p)\mathbf{R}_x^\epsilon\mathbf{a}(\theta_p)]}{\mathbf{a}'^H(\theta_p)\mathbf{R}_x^\epsilon\mathbf{a}'(\theta_p) + \text{Re}[\mathbf{a}^H(\theta_p)\mathbf{R}_x^\epsilon\mathbf{a}''(\theta_p)]}, \quad (47)$$

with $\mathbf{a}'(\theta_p) \stackrel{\text{def}}{=} \frac{d\mathbf{a}^H(\theta_p)}{d\theta_p}$ and $\mathbf{a}''(\theta_p) \stackrel{\text{def}}{=} \frac{d^2\mathbf{a}^H(\theta_p)}{d\theta_p^2}$.

Following the methodology of Subsection III-A2, the additional bias for finite value of N , that is of order $1/N$ can be derived, which gives

$$\text{E}(\hat{\theta}_{p,N}) - \theta_p = \text{AsBias}(\theta_p) + \frac{b_p}{N} + o\left(\frac{1}{N}\right),$$

see e.g., the involved expression of b_p for the Capon algorithm [91, rel. (35)].

In the same way, the covariance $\text{E}[(\hat{\theta}_N - \text{E}(\hat{\theta}_N))(\hat{\theta}_N - \text{E}(\hat{\theta}_N))^T]$ which is of order $1/N$ can be derived. It is obtained with $\bar{\boldsymbol{\theta}} \stackrel{\text{def}}{=} [\bar{\theta}_1, \dots, \bar{\theta}_P]^T$

$$\text{E}[(\hat{\theta}_N - \text{E}(\hat{\theta}_N))(\hat{\theta}_N - \text{E}(\hat{\theta}_N))^T] = \text{E}[(\hat{\theta}_N - \bar{\boldsymbol{\theta}})(\hat{\theta}_N - \bar{\boldsymbol{\theta}})^T] + o\left(\frac{1}{N}\right) = \frac{\mathbf{R}_\theta}{N} + o\left(\frac{1}{N}\right),$$

see e.g., the involved expression [42, rel. (24)] of \mathbf{R}_θ associated with a source for several parameters per source. The relative values of the asymptotic bias, additional bias and standard deviation depend on the SNR, M and N , but in practice the standard deviation is typically dominant over the asymptotic bias and additional bias (see examples given in [91]).

Finally, note that in the particular case of a single source, uniform white noise ($\mathbf{R}_n = \sigma_n^2\mathbf{I}$) and an arbitrary number d of parameters of the source (here $\boldsymbol{\theta} = (\theta_1, \dots, \theta_d)^T$), it has been proved [38], that $\hat{\theta}_N$ given by these two beamforming-based algorithms is asymptotically unbiased (AsBias(θ_p) given by (47) is zero), if and only if $\|\mathbf{a}(\boldsymbol{\theta})\|$ is constant. Furthermore, based on the general expressions (48) of the FIM⁶

$$\mathbf{FIM}(\boldsymbol{\theta}) = \frac{2N\sigma_s^4}{\sigma_n^2(\sigma_n^2 + \|\mathbf{a}(\boldsymbol{\theta})\|^2\sigma_s^2)} \text{Re} \left[\|\mathbf{a}(\boldsymbol{\theta})\|^2 \mathbf{D}(\boldsymbol{\theta})^H \mathbf{D}(\boldsymbol{\theta}) - \mathbf{D}(\boldsymbol{\theta})^H \mathbf{a}(\boldsymbol{\theta}) \mathbf{a}^H(\boldsymbol{\theta}) \mathbf{D}(\boldsymbol{\theta}) \right], \quad (48)$$

where $\mathbf{D}(\boldsymbol{\theta})$ is defined here by $[\partial\mathbf{a}(\boldsymbol{\theta})/\partial\theta_1, \dots, \partial\mathbf{a}(\boldsymbol{\theta})/\partial\theta_d]$, for d parameters associated with a single source, and expression [42, rel. (24)] of \mathbf{R}_θ specialized to $\mathbf{R}_n = \sigma_n^2\mathbf{I}$, it has been proved that $\frac{1}{N}\mathbf{R}_\theta = \mathbf{FIM}^{-1}(\boldsymbol{\theta})$, i.e., the conventional and Capon algorithms are asymptotically efficient, if and only if $\|\mathbf{a}(\boldsymbol{\theta})\|$ is constant.

B. Maximum likelihood algorithms

1) *Stochastic and deterministic ML algorithms:* As discussed in Subsection II-B, the two main models for the sensor array problem in Gaussian noise, corresponding to stochastic and deterministic modeling of the source signals lead to two different Gaussian distributions of the measurements $\{\mathbf{x}(t)\}_{t_1, \dots, t_N}$, and consequently to two different log-likelihoods $l(\boldsymbol{\alpha}) = \ln p(\mathbf{x}; \boldsymbol{\alpha})$, where the unknown parameter $\boldsymbol{\alpha}$ is respectively given by (4) and (35).

With some algebraic effort, the stochastic ML criterion $l(\boldsymbol{\alpha})$ can be concentrated w.r.t. \mathbf{R}_s and σ_n^2 (see e.g., [43],[84]), thus reducing the dimension of the required numerical maximization to the required P DOAs ($\theta_1, \dots, \theta_P$) and giving the following optimization problem:

$$\hat{\boldsymbol{\theta}}_N^{\text{SML}} = \arg \min_{\boldsymbol{\theta} \in \Theta^P} J_{\text{SML}}[\boldsymbol{\theta}, \mathbf{R}_{x,N}], \quad (49)$$

with

$$J_{\text{SML}}[\boldsymbol{\theta}, \mathbf{R}_{x,N}] = \ln[\det(\mathbf{A}(\boldsymbol{\theta})\mathbf{R}_{s,N}(\boldsymbol{\theta})\mathbf{A}^H(\boldsymbol{\theta}) + \sigma_{n,N}^2(\boldsymbol{\theta})\mathbf{I})], \quad (50)$$

where

$$\mathbf{R}_{s,N}(\boldsymbol{\theta}) = \mathbf{A}^\#(\boldsymbol{\theta})[\mathbf{R}_{x,N} - \sigma_{n,N}^2(\boldsymbol{\theta})\mathbf{I}]\mathbf{A}^{\#H}(\boldsymbol{\theta}) \quad \text{and} \quad \sigma_{n,N}^2(\boldsymbol{\theta}) = \frac{1}{M-P} \text{Tr}[\boldsymbol{\Pi}_{\mathbf{A}}^\perp(\boldsymbol{\theta})\mathbf{R}_{x,N}], \quad (51)$$

⁶For one parameter ($d = 1$) or $\|\mathbf{a}(\boldsymbol{\theta})\|$ constant, (48) can be simplified by withdrawing the real operator [67, rel. (49)].

where $\Pi_{\mathbf{A}}^{\perp}(\boldsymbol{\theta}) = \mathbf{I} - \mathbf{A}(\boldsymbol{\theta})\mathbf{A}^{\#}(\boldsymbol{\theta})$ is the orthogonal projector onto the null space of \mathbf{A}^H . Despite its reduction of the parameter space, $J_{\text{SML}}[\boldsymbol{\theta}, \mathbf{R}_{x,N}]$ is a complicated nonlinear expression in $\boldsymbol{\theta}$, that cannot be analytically minimized. Consequently, numerical optimization procedures are required.

Remark that in this modeling, the obvious a priori information that \mathbf{R}_s is positive semi-definite has not been taken into account. This knowledge, and more generally, the prior that \mathbf{R}_s is positive semi-definite of rank r smaller or equal than P can be included in the modeling by the parametrization $\mathbf{R}_s = \mathbf{L}\mathbf{L}^H$, where \mathbf{L} is a $P \times r$ lower triangular matrix. But this modification will have no effect for "large enough N " since $\widehat{\mathbf{R}}_s$ given by (51) is a weakly consistent estimate of \mathbf{R}_s [64]. And since this new parametrization leads to significantly more involved optimization, the unrestricted parametrization of \mathbf{R}_s used in (50) appears to be preferable.

Due to the quadratic dependence of the deterministic ML criterion $l(\boldsymbol{\alpha})$ in the parameters $\{\mathbf{s}(t)\}_{t_1, \dots, t_N}$, its concentration w.r.t. $\{\mathbf{s}(t)\}_{t_1, \dots, t_N}$ and σ_n^2 is much more simpler than for the stochastic ML criterion. It gives the following new ML estimator

$$\widehat{\boldsymbol{\theta}}_N^{\text{DML}} = \arg \min_{\boldsymbol{\theta} \in \Theta^P} J_{\text{DML}}[\boldsymbol{\theta}, \mathbf{R}_{x,N}], \quad (52)$$

with

$$J_{\text{DML}}[\boldsymbol{\theta}, \mathbf{R}_{x,N}] = \text{Tr}[\Pi_{\mathbf{A}}^{\perp}(\boldsymbol{\theta})\mathbf{R}_{x,N}]. \quad (53)$$

Comparing (53) and (50), we see that the dependence in $\boldsymbol{\theta}$ of the DML criterion is simpler than for the SML criterion. But both criteria require nonlinear P th-dimensionnal minimizations with a large number of local minima that give two different estimates $\boldsymbol{\theta}$, except for a single source for which the minimization of (53) and (50) reduce to the maximization of the common criteria

$$\frac{\mathbf{a}^H(\boldsymbol{\theta})\mathbf{R}_{x,N}\mathbf{a}(\boldsymbol{\theta})}{\|\mathbf{a}(\boldsymbol{\theta})\|^2}.$$

This implies that when the norm of the steering vector $\mathbf{a}(\boldsymbol{\theta})$ is constant (which is generally assumed), the conventional and Capon beamforming, SML and DML algorithms coincide and thus conventional and Capon beamforming and DML algorithms inherit the asymptotical efficiency of the SML algorithm. Note that this property extends to several parameters per source.

2) *Asymptotic properties of ML algorithms:* We consider in this Subsection, the asymptotic properties of DML or SML algorithms used under the respectively, deterministic and circular Gaussian stochastic modeling of the sources. In the field of asymptotic performance characterization of DML or SML algorithms, asymptotic generally refers to either the number N of snapshots or the SNR value.

First, consider the asymptotic properties w.r.t. N , that are the most known. Under regularity conditions that are satisfied by the SML algorithm, the general properties of ML estimation states that $\widehat{\boldsymbol{\theta}}_N^{\text{SML}}$ is consistent and asymptotically efficient and Gaussian distributed, more precisely

$$\sqrt{N} (\widehat{\boldsymbol{\theta}}_N^{\text{SML}} - \boldsymbol{\theta}) \xrightarrow{\mathcal{L}} \mathcal{N}_R(\mathbf{0}; \mathbf{R}_{\boldsymbol{\theta}}^{\text{SML}}) \quad \text{with} \quad \mathbf{R}_{\boldsymbol{\theta}}^{\text{SML}} = N\text{CRB}_{\text{CG}}(\boldsymbol{\theta}), \quad (54)$$

where $\text{CRB}_{\text{CG}}(\boldsymbol{\theta})$ is given by (33). This property of the SML algorithm extends to nonuniform white and unknown parameterized noise field in [65] and [39], respectively, and to general noncircular Gaussian stochastic modeling of the sources with the associated $\text{CRB}_{\text{NCG}}(\boldsymbol{\theta})$ [22], [2]. Note that to circumvent the difficulty to extract the " $\boldsymbol{\theta}$ corner" from the inverse of $\mathbf{FIM}(\boldsymbol{\alpha})$, a matrix closed-form expression of $\text{CRB}_{\text{CG}}(\boldsymbol{\theta})$ has been first obtained in an indirect manner by an asymptotic analysis of the SML estimator [81], [63]. Then, only ten years later, this CRB has been obtained directly from the extended Slepian-Bangs formula [87], [39].

As for the DML algorithm, since the signal waveforms themselves are regarded as unknown parameters, it follows that the number of unknown parameters $\boldsymbol{\alpha}$ (35) in the modeling, grows without limit with increasing N , the general asymptotic properties of the ML no longer apply. More precisely, the DML estimate of $\boldsymbol{\theta}$ is weakly consistent, whereas the DML estimate of $\{\mathbf{s}(t_n)\}_{n=1, \dots, N}$ is inconsistent. The asymptotic distribution of $\widehat{\boldsymbol{\theta}}_N^{\text{DML}}$ has been derived in [82], [93]

$$\sqrt{N} (\widehat{\boldsymbol{\theta}}_N^{\text{DML}} - \boldsymbol{\theta}) \xrightarrow{\mathcal{L}} \mathcal{N}_R(\mathbf{0}; \mathbf{R}_{\boldsymbol{\theta}}^{\text{DML}}) \quad (55)$$

with

$$\mathbf{R}_\theta^{\text{DML}} = N\text{CRB}_{\text{Det}}(\boldsymbol{\theta}) + 2N^2\text{CRB}_{\text{Det}}(\boldsymbol{\theta})\text{Re}[(\mathbf{D}^H\boldsymbol{\Pi}_x\mathbf{D}) \odot (\mathbf{A}^H\mathbf{A})^{-T}] \text{CRB}_{\text{Det}}(\boldsymbol{\theta}), \quad (56)$$

where $\text{CRB}_{\text{Det}}(\boldsymbol{\theta})$ is given by (36). Note that the inequality $\frac{1}{N}\mathbf{R}_\theta^{\text{DML}} \leq \text{CRB}_{\text{Det}}(\boldsymbol{\theta})$ in (56) does not follow from the Cramer-Rao inequality theory directly, because the Cramer-Rao inequality requires that the number of unknown parameters be finite. As the number of real-valued parameters in $\boldsymbol{\alpha}$ (35) is $P + 2NP + 1$, it increases with N and the Cramer-Rao inequality does not apply here. Note that the DML estimates of $\{\mathbf{s}(t_n)\}_{n=1,\dots,N}$ are indeed asymptotically unbiased, despite being non-consistent.

Furthermore, it has been proved in [82], that if the DML algorithm is used under the circular Gaussian stochastic modeling of the sources, the asymptotic distribution (54) of $\hat{\boldsymbol{\theta}}_N^{\text{DML}}$ is preserved. But under this assumption on the sources, the DML algorithm is suboptimal, and thus $\frac{1}{N}\mathbf{R}_\theta^{\text{DML}} \geq \text{CRB}_{\text{CG}}(\boldsymbol{\theta})$. Finally comparing directly the expressions (33) and (36) of the Cramer-Rao bound by applying the matrix inversion lemma, it is straightforward to prove that $\text{CRB}_{\text{CG}}(\boldsymbol{\theta}) \geq \text{CRB}_{\text{Det}}(\boldsymbol{\theta})$. This allows one to relate $\mathbf{R}_\theta^{\text{DML}}$, $\mathbf{R}_\theta^{\text{SML}}$, $\text{CRB}_{\text{CG}}(\boldsymbol{\theta})$ and $\text{CRB}_{\text{Det}}(\boldsymbol{\theta})$ by the following relation:

$$\frac{1}{N}\mathbf{R}_\theta^{\text{DML}} \geq \frac{1}{N}\mathbf{R}_\theta^{\text{SML}} = \text{CRB}_{\text{CG}}(\boldsymbol{\theta}) \geq \text{CRB}_{\text{Det}}(\boldsymbol{\theta}). \quad (57)$$

In particular, for a single source with q parameters, we have

$$\text{CRB}_{\text{CG}}(\boldsymbol{\theta}) = \left(1 + \frac{\sigma_n^2}{\|\mathbf{a}(\boldsymbol{\theta})\|^2\sigma_s^2}\right) \text{CRB}_{\text{Det}}(\boldsymbol{\theta}), \quad (58)$$

with $\text{CRB}_{\text{CG}}(\boldsymbol{\theta}) = \mathbf{FIM}^{-1}(\boldsymbol{\theta})$, where $\mathbf{FIM}(\boldsymbol{\theta})$ is given by (48).

Finally, note an asymptotic robustness property [81], [63] of the SML and DML algorithms that states that the asymptotic distribution of $\hat{\boldsymbol{\theta}}_N^{\text{SML}}$ and $\hat{\boldsymbol{\theta}}_N^{\text{DML}}$ is preserved whatever the modeling of the source: circular Gaussian distributed with $\text{E}[\mathbf{s}(t)\mathbf{s}^H(t)] = \mathbf{R}_s$ or modeled by arbitrary second-order ergodic signals with $\mathbf{R}_s = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=1}^N \mathbf{s}(t_n)\mathbf{s}^H(t_n)$. We will present a more general asymptotic robustness property that applies to a large category of second-order algorithms in Subsection IV-C. The fact that the SML algorithm always outperforms (for $P > 1$) the DML algorithm, provides strong justifications for the appropriateness of the stochastic modeling of sources for the DOA estimation problem.

Consider now, the asymptotic properties of the SML and DML algorithms w.r.t. SNR, used under their respective source model assumptions. It has been proved in [71], that under the circular Gaussian assumption of the sources, the SML estimates $\hat{\boldsymbol{\theta}}_N^{\text{SML}}$ is asymptotically (w.r.t. SNR) non-Gaussian distributed and non-efficient, i.e., $\tilde{\boldsymbol{\theta}}_{\sigma_n} \stackrel{\text{def}}{=} \frac{1}{\sigma_n}(\hat{\boldsymbol{\theta}}_N^{\text{SML}} - \boldsymbol{\theta})$ converges in distribution to a non-Gaussian distribution, when σ_n tends to zero, with N fixed, with $\lim_{\sigma_n \rightarrow 0} \text{E}[\tilde{\boldsymbol{\theta}}_{\sigma_n}\tilde{\boldsymbol{\theta}}_{\sigma_n}^T] \geq \lim_{\sigma_n \rightarrow 0} \frac{1}{\sigma_n^2}\text{CRB}_{\text{CG}}(\boldsymbol{\theta})$. In practice, $\hat{\boldsymbol{\theta}}_N^{\text{SML}}$ is non-Gaussian distributed and non-efficient at high SNR, only for a very small number N of snapshots⁷. For example, for a single source, using (37), it is proved in [71] that

$$\lim_{\sigma_n \rightarrow 0} \text{E}[\tilde{\boldsymbol{\theta}}_{\sigma_n}\tilde{\boldsymbol{\theta}}_{\sigma_n}^T] = \frac{N}{N-1} \lim_{\sigma_n \rightarrow 0} \frac{1}{\sigma_n^2}\text{CRB}_{\text{CG}}(\boldsymbol{\theta}) = \frac{N}{N-1} \left(\frac{1}{Nh_1\sigma_1^2}\right),$$

(see (34) for the second equality), where h_1 is defined just after (34). These properties contrast with the DML algorithm used under the deterministic modeling of the sources, which is proved [70] to be asymptotically (w.r.t. SNR) Gaussian distributed and efficient, i.e., $\frac{1}{\sigma_n}(\hat{\boldsymbol{\theta}}_N^{\text{DML}} - \boldsymbol{\theta}) \xrightarrow{\mathcal{L}} \mathcal{N}_R\left(\mathbf{0}; \frac{1}{2N} \{\text{Re}[(\mathbf{D}^H\boldsymbol{\Pi}_x\mathbf{D}) \odot \mathbf{R}_s]\}^{-1}\right)$ when σ_n tends to zero, with N arbitrary fixed. These results are consistent with those of [81]. In practice for very high SNR and "not too small" N , (57) becomes

$$\frac{1}{N}\mathbf{R}_\theta^{\text{DML}} \approx \frac{1}{N}\mathbf{R}_\theta^{\text{SML}} = \text{CRB}_{\text{CG}}(\boldsymbol{\theta}) \approx \text{CRB}_{\text{Det}}(\boldsymbol{\theta}). \quad (59)$$

⁷In practice the approximate covariances deduced from the asymptotic analysis w.r.t. the number of snapshots are also valid for high SNR with fixed "not too small number" of snapshots for the second-order DOA algorithms. But note that there is no theoretical result on the asymptotic distribution of the sample projector w.r.t. the SNR.

Furthermore, it has been proved in [81], that (59) is also valid for $M \gg 1$. The asymptotic distribution of the DOA estimate w.r.t. M (for finite data) of the SML and DML algorithms has been studied in [96]. The strong consistency has been proved for both ML algorithms. Furthermore, unlike the previously studied large sample case, the asymptotic covariance matrices of the DOA estimates coincide with the deterministic CRB (36) for the SML and DML algorithms. The asymptotic distribution of the DOA estimates given by subspace-based algorithms has been studied in [92], when $M, N \rightarrow \infty$, whereas M/N converges to a strictly positive constant. In this asymptotic regime, it is proved, in particular, that these traditional DOA estimates are not consistent. The threshold and the so-called subspace swap of the SML and MUSIC algorithms have been studied w.r.t. N , M and SNR (see e.g., [45]). Furthermore, a new consistent subspace-based estimate has been proposed, which outperforms the standard subspace-based methods for values of M and N of the same order of magnitude [92].

3) *Large sample ML approximations*: Since the SML and DML algorithms are often deemed exceedingly complex, suboptimal algorithms are of interest. Many such algorithms have been proposed in the literature and surprisingly, some of them are asymptotically as accurate as the ML algorithms, but with a reduced computational cost. These algorithms have been derived, either by approximations of the ML criteria by neglecting terms that do not affect the asymptotic properties of the estimates, or by using a purely geometrical point of view. We present this latter approach that allows one to unify a large number of algorithms [64]. These algorithms rely on the geometrical properties of the spectral decomposition of the covariance matrix \mathbf{R}_x :

$$\mathbf{R}_x = \mathbf{E}_s \mathbf{\Lambda}_s \mathbf{E}_s^H + \sigma_n^2 \mathbf{E}_n \mathbf{E}_n^H$$

with $\mathbf{E}_s = [\mathbf{e}_1, \dots, \mathbf{e}_r]$, $\mathbf{\Lambda}_s = \text{Diag}(\lambda_1, \dots, \lambda_r)$ and $\mathbf{E}_n = [\mathbf{e}_{r+1}, \dots, \mathbf{e}_M]$ where r is the rank of \mathbf{R}_s , associated with the consistent estimates

$$\mathbf{R}_{x,N} \stackrel{\text{def}}{=} \frac{1}{N} \sum_{n=1}^N \mathbf{x}(t_n) \mathbf{x}^H(t_n) = \mathbf{E}_{s,N} \mathbf{\Lambda}_{s,N} \mathbf{E}_{s,N}^H + \sigma_{n,N}^2 \mathbf{E}_{n,N} \mathbf{E}_{n,N}^H. \quad (60)$$

These algorithms can be classified as signal subspace-based and noise subspace-based fitting algorithms. The former algorithms based on $\text{Span}(\mathbf{E}_s) \subseteq \text{Span}(\mathbf{A}(\boldsymbol{\theta}))$ are given by the following optimization:

$$\hat{\boldsymbol{\theta}}_N^{\text{SSF}} = \arg \min_{\boldsymbol{\theta} \in \Theta^P} \text{Tr}[\mathbf{\Pi}_{\mathbf{A}}^\perp(\boldsymbol{\theta}) \mathbf{E}_{s,N} \mathbf{W} \mathbf{E}_{s,N}^H], \quad (61)$$

where \mathbf{W} is a weighting $r \times r$ positive definite matrix to be specified. And the latter algorithms based on $\mathbf{E}_n^H \mathbf{A}(\boldsymbol{\theta}) = \mathbf{0}$, that is valid only if the source covariance matrix is nonsingular ($r = P$), are given by

$$\hat{\boldsymbol{\theta}}_N^{\text{NSF}} = \arg \min_{\boldsymbol{\theta} \in \Theta^P} \text{Tr}[\mathbf{U} \mathbf{A}^H(\boldsymbol{\theta}) \mathbf{E}_{n,N} \mathbf{E}_{n,N}^H \mathbf{A}(\boldsymbol{\theta})], \quad (62)$$

where \mathbf{U} is a weighting $P \times P$ positive definite matrix to be specified.

Introduced from a purely geometrical point of view, these two classes of algorithms present unexpected relations with the previously described ML algorithms. First, for arbitrary positive definite weighting matrices \mathbf{W} and \mathbf{U} , the estimates $\hat{\boldsymbol{\theta}}_N^{\text{SSF}}$ and $\hat{\boldsymbol{\theta}}_N^{\text{NSF}}$ given respectively by (61) and (62), are weakly consistent. Second, for the weighting matrices that give the lowest covariance matrix of the asymptotic distribution of $\hat{\boldsymbol{\theta}}_N^{\text{SSF}}$ and $\hat{\boldsymbol{\theta}}_N^{\text{NSF}}$, that are respectively given [64] by

$$\mathbf{W}_{\text{opt}} = (\mathbf{\Lambda}_s - \sigma_n^2 \mathbf{I})^2 \mathbf{\Lambda}_s^{-1} \quad \text{and} \quad \mathbf{U}_{\text{opt}} = \mathbf{A}^\#(\boldsymbol{\theta}_0) \mathbf{E}_s \mathbf{W}_{\text{opt}} \mathbf{E}_s^H \mathbf{A}^{\#H}(\boldsymbol{\theta}_0),$$

where $\boldsymbol{\theta}_0$ denotes here the true value of the DOAs, the associated estimates $\hat{\boldsymbol{\theta}}_N^{\text{SSF}}$ and $\hat{\boldsymbol{\theta}}_N^{\text{NSF}}$ are asymptotically equivalent to $\hat{\boldsymbol{\theta}}_N^{\text{SML}}$ (i.e., $\sqrt{N}(\hat{\boldsymbol{\theta}}_N^{\text{SSF}} - \hat{\boldsymbol{\theta}}_N^{\text{SML}}) \rightarrow \mathbf{0}$ and $\sqrt{N}(\hat{\boldsymbol{\theta}}_N^{\text{NSF}} - \hat{\boldsymbol{\theta}}_N^{\text{SML}}) \rightarrow \mathbf{0}$ in probability as $N \rightarrow \infty$) and thus have the same asymptotic distribution that the SML algorithm. Furthermore and fortunately, this property extends for any weakly consistent estimates \mathbf{W}_N and \mathbf{U}_N of respectively \mathbf{W}_{opt} and \mathbf{U}_{opt} , e.g., derived from the spectral decomposition of the sample covariance matrix $\mathbf{R}_{x,N}$ (60) with $\sigma_{n,N}^2$ is the average of $M - r$ smallest eigenvalues of $\mathbf{R}_{x,N}$ and with $\boldsymbol{\theta}_0$ is replaced by a weakly consistent estimates of $\boldsymbol{\theta}$. This implies a two steps procedure to run the optimal noise subspace-based fitting algorithm. Due to this drawback, the signal subspace-based fitting algorithm with the weighting $\mathbf{W}_N = (\mathbf{\Lambda}_{s,N} - \sigma_{n,N}^2 \mathbf{I})^2 \mathbf{\Lambda}_{s,N}^{-1}$, denoted weighted subspace

fitting (WSF) algorithm, is preferred to the noise subspace-based fitting algorithms.

Finally, note that this algorithm is based on eigenvalues and eigenvectors of the sample covariance matrix $\mathbf{R}_{x,N}$. This contrasts with the subspace-based algorithms whose asymptotic statistical properties will be studied in Subsection IV-D that are based on the noise or signal orthogonal projector $\mathbf{\Pi}_{x,N}$ associated with $\mathbf{R}_{x,N}$ only. Note that general properties of subspace-based estimators focused on asymptotic invariance of these estimators have been given in [12].

C. Second-order algorithms

Most of the narrowband DOA algorithms presented in the literature are second-order algorithms, i.e., are based on the sample covariance $\mathbf{R}_{x,N} \stackrel{\text{def}}{=} \frac{1}{N} \sum_{n=1}^N \mathbf{x}(t_n) \mathbf{x}^H(t_n)$ or more generally on $\mathbf{R}_{\tilde{x},N} \stackrel{\text{def}}{=} \frac{1}{N} \sum_{n=1}^N \tilde{\mathbf{x}}(t_n) \tilde{\mathbf{x}}^H(t_n)$. To prove common properties of this class of algorithm, it is useful to use the functional analysis presented in Subsection III-A1

$$\{\mathbf{x}(t)\}_{t_1, \dots, t_N} \mapsto \mathbf{R}_{x,N} \xrightarrow{\text{alg}} \hat{\boldsymbol{\theta}}_N, \quad (63)$$

in which any second-order algorithm is a mapping **alg** that generally satisfies

$$\mathbf{alg}(\mathbf{A}(\boldsymbol{\theta}) \mathbf{R}_s \mathbf{A}^H(\boldsymbol{\theta}) + \sigma_n^2 \mathbf{I}) = \boldsymbol{\theta} \quad \text{for any } \boldsymbol{\theta} \in \Theta^P, \quad (64)$$

but not necessarily for all $P \times P$ Hermitian positive semi-definite matrix \mathbf{R}_s . Depending on the a priori knowledge about \mathbf{R}_s , that is required by the second-order algorithms **alg**, different constraints are satisfied by the \mathbb{C} -differential matrix $\mathbf{D}_{R_x, \theta}^{\text{alg}}$ of the algorithm at the point \mathbf{R}_x (22). In particular, it has been proved the following main two constraints [19]:

$$\mathbf{D}_{R_x, \theta}^{\text{alg}}(\mathbf{A}(\boldsymbol{\theta}) \otimes \mathbf{A}(\boldsymbol{\theta})) = \mathbf{0} \quad \text{for } \mathbf{R}_s \text{ unstructured} \quad (65)$$

$$\mathbf{D}_{R_x, \theta}^{\text{alg}}(\mathbf{a}(\theta_p) \otimes \mathbf{a}(\theta_p)) = \mathbf{0}, p = 1, \dots, P \quad \text{for } \mathbf{R}_s \text{ structured diagonal.} \quad (66)$$

Using these constraints, the general expression \mathbf{R}_{R_x} of the covariance of the asymptotic distribution of the sample covariance $\mathbf{R}_{x,N}$ [18] obtained under mild conditions for non independent measurements with arbitrary distributed sources and noise of finite fourth-order moments, and the general relation (23), that links \mathbf{R}_{R_x} and $\mathbf{D}_{R_x, \theta}^{\text{alg}}$ to the covariance \mathbf{R}_θ of the asymptotic distribution of $\hat{\boldsymbol{\theta}}_N$, allows one to prove the following two results, that extend a robustness property presented in [10]:

- For any second-order algorithms based on $\mathbf{R}_{x,N}$, that do not require the sources spatially uncorrelated and when the noise signals $\{\mathbf{n}(t)\}_{t_1, \dots, t_N}$ are temporally uncorrelated, \mathbf{R}_θ is invariant to the distribution, the second-order noncircularity and the temporal distribution of the sources, but depends on the distribution of the noise through its second-order and fourth-order moments. In particular for circular Gaussian noise, the asymptotic distribution of $\hat{\boldsymbol{\theta}}_N$ are those of the standard complex circular Gaussian case.
- For any second-order algorithms based on $\mathbf{R}_{x,N}$ that require the sources spatially uncorrelated and/or when the noise signals $\{\mathbf{n}(t)\}_{t_1, \dots, t_N}$ are temporally correlated, \mathbf{R}_θ is sensitive to the distribution, the second-order noncircularity and the temporal distribution of the sources.

Note that the majority of the second-order algorithms (e.g., the beamforming, ML, MUSIC, Min Norm, ESPRIT algorithms) does not require spatially uncorrelated sources. In contrast, second-order techniques based on state-space realizations (e.g., the Toeplitz approximation method (TAM), see [57]) and Toeplitzation or augmentation with ULA or uniform rectangular arrays, require this uncorrelation, and thus the asymptotic distribution of $\hat{\boldsymbol{\theta}}_N$ will be generally (except for a single source, for which the constraint (66) reduces to (65)) sensitive to the distribution, the second-order noncircularity or the temporal distribution of the sources, even when the noise is temporally uncorrelated.

To illustrate this sensitivity to the source distribution when the noise is temporally uncorrelated, we consider in Fig.2, the case of two equipowered and spatially uncorrelated sources impinging on a ULA of 10 sensors, $\theta_1 = 20^\circ$ and $\theta_2 = 30^\circ$, where the DOAs are estimated by the standard MUSIC algorithm after Toeplitzation. The sources are either white Gaussian, ARMA Gaussian (generated by a (10,10) Butterworth filter driven by a white circular Gaussian noise, where the bandwidth is fixed to 0.5) or harmonic. The centered frequencies of the ARMA and the frequencies of the harmonics are -0.25 and 0.25 . Fig.2 shows that the Toeplitzation

improves the performance for very weak SNR only, whereas is very sensitive to the distribution of the sources for high SNR.

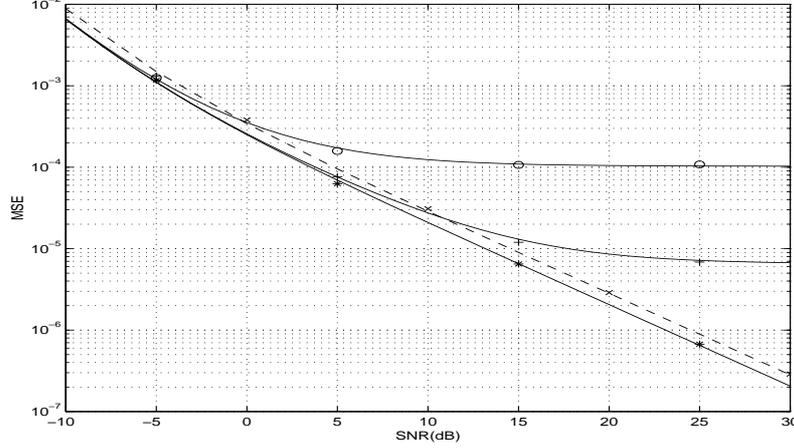


Fig.2 Theoretical and estimated MSE (with 500 Monte Carlo runs) of θ_1 versus the SNR, for respectively white (\circ), colored ($+$) and harmonic ($*$) signals for $N = 100$ after Toeplitzization (---) and without Toeplitzization (- - -).

Usually, performance analyses are evaluated as a function of the number N of observed snapshots without taking the sampling rate into account. In fact, depending on the value of this sampling rate, the collected samples $\mathbf{x}(t_n)$ are more or less temporally correlated and performance is affected. Thus, the interesting question arises as to how the asymptotic covariance of the DOA estimators, (denoted here $\hat{\boldsymbol{\theta}}_T$) varies with this sampling rate $\frac{1}{T_s}$ for a fixed observation interval $T = NT_s$. This question has been investigated in [19], in which the continuous-time noise envelope $\mathbf{n}(t)$ is spatially white and temporally white in the bandwidth $[-\frac{B}{2}, +\frac{B}{2}]$. It has been proved:

- If the signals $\mathbf{x}(t)$ are oversampled ($\frac{1}{T_s} > B$)

$$E[(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta})^T] \approx \frac{1}{BT} \mathbf{R}_\theta > \frac{1}{N} \mathbf{R}_\theta \quad \text{for } N \gg 1,$$

irrespective of the sample rate $1/T_s$.

- If the signals $\mathbf{x}(t)$ are subsampled ($\frac{1}{T_s} < B$)

$$E[(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}}_T - \boldsymbol{\theta})^T] \approx \frac{T_s}{T} \mathbf{R}_\theta = \frac{1}{N} \mathbf{R}_\theta > \frac{1}{BT} \mathbf{R}_\theta \quad \text{for } N \gg 1 \text{ and } BT_s \gg 1.$$

Consequently the array must be temporally oversampled, and the parameter of interest that characterizes performance ought not to be the number N of snapshots, but rather the observation interval T .

D. Subspace-based algorithms

We concentrate now on the family of second-order algorithms based on the orthogonal noise⁸ projector $\boldsymbol{\Pi}_{x,N}$ (10). These algorithms estimate $\boldsymbol{\theta}$, either by extrema-searching approaches (MUSIC, Min-Norm..), by polynomial rooting approaches (Pisarenko, root MUSIC and root Min-Norm for ULA), or by matrix shifting approaches (ESPRIT, TAM, Matrix pencil method). The most celebrated of these algorithms is the MUSIC algorithm, where $\boldsymbol{\theta}$ is estimated as the P deepest minima in a d -dimensional (for d parameters per source) of the following localisation function $J_{\text{MUSIC}}[\boldsymbol{\theta}, \boldsymbol{\Pi}_{x,N}]$

$$J_{\text{MUSIC}}[\boldsymbol{\theta}, \boldsymbol{\Pi}_{x,N}] = \mathbf{a}^H(\boldsymbol{\theta}) \boldsymbol{\Pi}_{x,N} \mathbf{a}(\boldsymbol{\theta}), \quad (67)$$

⁸Note that since $\boldsymbol{\Pi}_x + \boldsymbol{\Pi}_x^\perp = \mathbf{I}$ and $\boldsymbol{\Pi}_{x,N} + \boldsymbol{\Pi}_{x,N}^\perp = \mathbf{I}$, all algorithm based on the orthogonal signal projector comes down to an algorithm based on the orthogonal noise projector.

of the so-called spatial null spectrum (or equivalently as the P highest peaks (maxima) of its inverse). This algorithm has given a plethora of variants. For example, in the particular case of the ULA, this standard MUSIC algorithm have been favorably replaced by the root MUSIC algorithm. Using the general methodology presented in Subsection III-A2, the asymptotic distribution of $\hat{\boldsymbol{\theta}}_N$ given by any subspace-based algorithms **alg** is simply derived from the expression of the \mathbb{C} -differential matrix $\mathbf{D}_{\Pi_x, \theta}^{\text{alg}}$ of the mapping $\Pi_{x, N} \xrightarrow{\text{alg}} \hat{\boldsymbol{\theta}}_N$ evaluated at $\Pi_x(\boldsymbol{\theta})$. For example, for the standard MUSIC algorithm, $\mathbf{D}_{\Pi_x, \theta}^{\text{MUSIC}}$ is straightforwardly obtained from the first-order expansion of $\left(\frac{\partial J_{\text{MUSIC}}(\theta; \Pi_{x, N})}{\partial \theta}\right)_{\theta=\theta_p+\delta\theta_{p, N}} = 0$ that gives for one parameter per source

$$\mathbf{D}_{\Pi_x, \theta}^{\text{MUSIC}} = \begin{bmatrix} \mathbf{d}_1^T \\ \vdots \\ \mathbf{d}_P^T \end{bmatrix} \quad \text{with} \quad \mathbf{d}_p^T = -\frac{1}{h_p} \left((\mathbf{a}'^T(\theta_p) \otimes \mathbf{a}^H(\theta_p)) + (\mathbf{a}^T(\theta_p) \otimes \mathbf{a}'^H(\theta_p)) \right), \quad p = 1, \dots, P, \quad (68)$$

with $\mathbf{a}'(\theta_p) \stackrel{\text{def}}{=} \frac{d\mathbf{a}(\theta_p)}{d\theta_p}$ and $h_p \stackrel{\text{def}}{=} 2\mathbf{a}'^H(\theta_p)\Pi_x\mathbf{a}'(\theta_p)$. Using (68) with (16) and (23) allow one to directly prove that the sequences $\sqrt{N}(\hat{\boldsymbol{\theta}}_N - \boldsymbol{\theta})$ converges in distribution to the zero-mean Gaussian distribution of covariance matrix given elementwise by $(\mathbf{R}_{\theta}^{\text{MUSIC}})_{k, l} = \frac{2}{h_k h_l} \text{Re} \left((\mathbf{a}^H(\theta_l)\mathbf{U}\mathbf{a}(\theta_k)) (\mathbf{a}'^H(\theta_k)\Pi_x\mathbf{a}'(\theta_l)) \right)$ and compactly by

$$\mathbf{R}_{\theta}^{\text{MUSIC}} = 2(\mathbf{H} \odot \mathbf{I})^{-1} \text{Re} \left(\mathbf{H} \odot (\mathbf{A}^H \mathbf{U} \mathbf{A})^T \right) (\mathbf{H} \odot \mathbf{I})^{-1}, \quad (69)$$

where $(\mathbf{H})_{p, p} \stackrel{\text{def}}{=} h_p$ and \mathbf{U} has been defined in Subsection III-A2. Note that these expressions have been derived in [80] by much more involved derivations based on the asymptotic distribution of the eigenvectors of the sample covariance matrix $\mathbf{R}_{x, N}$. Finally, note that if the sample orthogonal noise projector $\Pi_{x, N}$ is replaced by an adaptive estimator $\Pi_{x, \gamma}$ of Π_x , where γ is the step-size of an arbitrary constant step-size recursive stochastic algorithm (see e.g., [16] and [17]), it has been proved in [16] that $\sqrt{\gamma}(\hat{\boldsymbol{\theta}}_{\gamma} - \boldsymbol{\theta})$ converges in distribution to the zero-mean Gaussian distribution of covariance matrix given also by $\mathbf{R}_{\theta}^{\text{MUSIC}}$, where $\hat{\boldsymbol{\theta}}_{\gamma}$ is an adaptive estimate of $\boldsymbol{\theta}$ given by the MUSIC algorithm based on the specific adaptive estimate $\Pi_{x, \gamma}$ of Π_x studied in [16].

Using a similar approach [34], it has been proved that the Root MUSIC algorithm associated with the ULA, presents the same asymptotic distribution, but slightly outperforms the standard MUSIC algorithm outside the asymptotic regime. This analysis has been extended to MUSIC-like algorithms applied to the orthogonal noise projectors $\Pi'_{x, N}$ [resp. $\Pi_{\tilde{x}, N}$] associated with the complementary sample covariance $\mathbf{C}_{x, N}$ [the augmented sample covariance $\mathbf{R}_{\tilde{x}, N}$] matrices for the DOA estimation of arbitrary noncircular [resp. rectilinear] sources [3]. Finally, note that with our general methodology, all the expressions of the covariance $\mathbf{R}_{\theta}^{\text{MUSIC}}$ can be straightforward extended for several parameter per source.

The expression of the covariance (69) of the asymptotic distribution of $\hat{\boldsymbol{\theta}}_N$ given the standard MUSIC algorithm has been analyzed in detail (see e.g., [67], [80]). In particular it has been proved that the MUSIC algorithm is asymptotically efficient for a single source, an arbitrary number of parameters per source and $\|\mathbf{a}(\theta_1)\|$ depending on θ_1 , e.g., for one parameter per source

$$\frac{1}{N} \mathbf{R}_{\theta_1}^{\text{MUSIC}} = \text{CRB}_{\text{CG}}(\theta_1) = \frac{1}{N} \left(\frac{1}{h_1} \left[\frac{\sigma_n^2}{\sigma_1^2} + \frac{1}{\|\mathbf{a}(\theta_1)\|^2} \frac{\sigma_n^4}{\sigma_1^4} \right] \right).$$

For several sources, the MUSIC algorithm is in general asymptotically inefficient, in particular for correlated sources for which the efficiency degrades when the correlation between the sources increases. The degradation of performances are considerable for highly correlated sources for any value of the SNRs. In contrast, for uncorrelated sources, the MUSIC algorithm is asymptotically efficient when σ_n^2 tends to zero, in the following sense $\lim_{\sigma_n^2 \rightarrow 0} \left[\frac{1}{N} \mathbf{R}_{\theta}^{\text{MUSIC}} \right] [\text{CRB}_{\text{CG}}(\boldsymbol{\theta})]^{-1} = \mathbf{I}$. So, in practice, for uncorrelated sources, the MUSIC algorithm is asymptotically efficient for high SNRs of all the sources.

It is of utmost importance to investigate in what region of N and SNR, the asymptotic theoretical results can predict actual performance. But unfortunately, only Monte Carlo simulations can specify this region. We illustrate in the following the SNR threshold region for the SML, DML and MUSIC algorithm.

Consider two zero-mean circular Gaussian sources impinging on an ULA (2) with $M = 6$ (for which the 3dB

bandwidth is about 8°) and a spatially uniform white noise (3). The source $s_1(t)$ consist of a strong direct path at $\theta_1 = 0^\circ$ relative to array broadside and a weaker (multipath at $\theta_2 = 4^\circ$ at -3dB w.r.t. $s_1(t)$). The correlation between $s_1(t)$ and $s_2(t)$ is 0.99 giving thus the source covariance matrix $\mathbf{R}_s = \begin{bmatrix} 1 & 0.7 \\ 0.7 & 0.5 \end{bmatrix}$. Fig.3 shows the root mean square error (RMSE) of the estimated DOA $\hat{\theta}_1$ by the MUSIC algorithm w.r.t. the SNR defined by σ_1^2/σ_n^2 , compared with the theoretical standard deviation (TSD) $\sqrt{\frac{1}{N}(\mathbf{R}_\theta^{\text{MUSIC}})_{1,1}}$ and the square root of the stochastic CRB $\sqrt{\text{CRB}_{\text{CG}}(\theta_1)}$. We see from this figure that the MUSIC algorithm is not efficient at all for highly correlated sources. Furthermore, the domain of validity of the asymptotic regime is here very limited, i.e., for $N = 1000$, $\text{SNR} > 30\text{dB}$ is required.

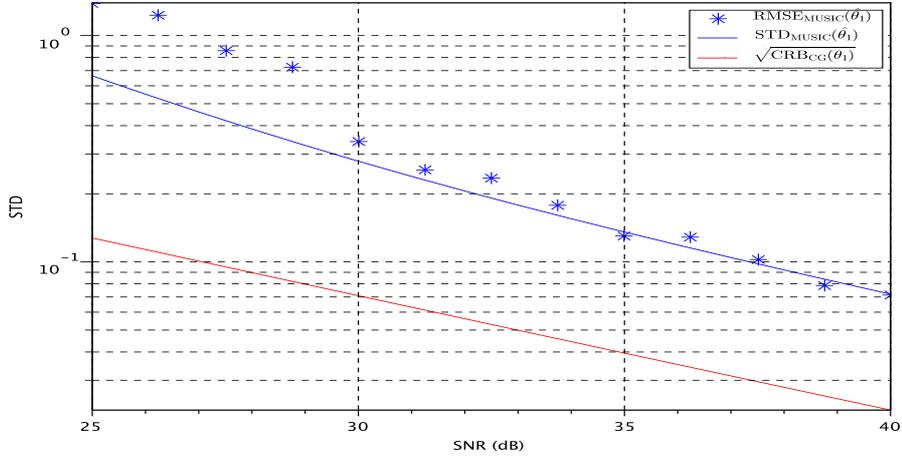


Fig.3 RMSE of $\hat{\theta}_1$ estimated by the MUSIC algorithm (averaged on 1000 runs) compared with the theoretical standard deviation and the square root of the stochastic CRB, as a function of the SNR for $N = 1000$.

With the same parameters, Fig.4 shows the RMSE of the estimated DOA $\hat{\theta}_1$ by the SML and DML algorithms which are compared with the TSD $\sqrt{\frac{1}{N}(\mathbf{R}_\theta^{\text{SML}})_{1,1}}$ and $\sqrt{\frac{1}{N}(\mathbf{R}_\theta^{\text{DML}})_{1,1}}$ and the square roots of the CRBs $\sqrt{\text{CRB}_{\text{CG}}(\theta_1)}$ and $\sqrt{\text{CRB}_{\text{DET}}(\theta_1)}$. We see from this figure that the numerical values of the four expressions of (57) are very close and the performance of the two ML algorithms are very similar except for the SNR threshold region for which the SML algorithm is efficient for $\text{SNR} > 0\text{dB}$ with $N = 1000$. Finally, comparing Fig.3 and Fig.4, we see that both ML algorithms largely outperform the MUSIC algorithm for highly correlated sources.

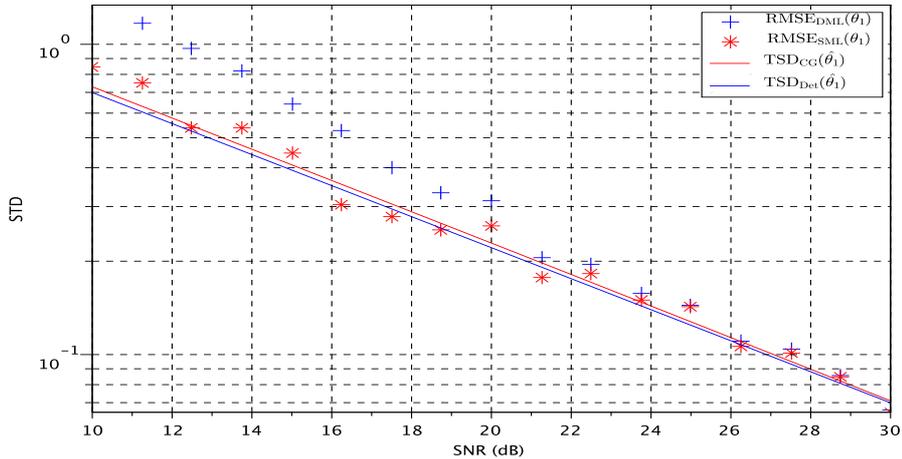


Fig.4 RMSE of $\hat{\theta}_1$ estimated by the SML and DML algorithms (averaged on 1000 runs) compared with the theoretical standard deviations and the square root of the stochastic and deterministic CRBs, as a function of the SNR for $N = 1000$.

E. Robustness of algorithms

We distinguish in this subsection, the robustness of the DOA estimation algorithms w.r.t. the narrowband assumption and to array modeling errors, because for the array modeling errors, the model (1) remains valid with a modified steering matrix, in contrast to the violation of narrowband assumption, for which (1) must be modified.

1) *Robustness w.r.t. the narrowband assumption:* As the wideband assumption generally requires an increased computational complexity compared to the narrowband ones, it is of interest to examine if the narrowband methods can be used for a sufficiently wide bandwidth without sacrificing performance. Some responses to this question have been given in [79] for symmetric spectra w.r.t. the demodulation frequency and in [20] for non-symmetric spectra and/or offset of the centered value of the spectra w.r.t. the demodulation frequency f_0 . In these assumptions, the model (1) of the complex envelope of the measurements becomes

$$\mathbf{x}(t) = \sum_{p=1}^P \int_{-B/2}^{+B/2} \mathbf{a}(\theta_p, f_0 + f) e^{i2\pi ft} d\mu_p(f) + \mathbf{n}(t), \quad (70)$$

where $\mathbf{a}(\theta_p, \nu) \stackrel{\text{def}}{=} [e^{i2\pi\nu\tau_{1,p}}, \dots, e^{i2\pi\nu\tau_{M,p}}]^T$ (with $\mathbf{a}(\theta_p, f_0) = \mathbf{a}(\theta_p)$) and $\mu_p(f)$ is the spectral measure of the p th source. Using the general methodology explained in Subsection III-A, based on a first-order expansion of the DOA estimate $\hat{\theta}_N = \mathbf{alg}(\mathbf{\Pi}_{x,N})$ in the neighborhood of $\mathbf{\Pi}_x$, (where $\mathbf{\Pi}_{x,N}$ and $\mathbf{\Pi}_x$ are the orthogonal projectors onto the noise subspace associated with the covariance of (70) and (1), respectively), general closed-form expressions of the asymptotic (w.r.t. the number of snapshots and source bandwidth) for arbitrary subspace-based algorithm have been derived in [20]. It is found that the behavior of these DOA estimators strongly depends on the symmetry of the source spectra w.r.t. their centered value and on the offset of this centered value w.r.t. f_0 . It is showed that the narrowband SOS-based algorithms are much more sensitive to the frequency offset than to the bandwidth.

In particular for source spectra $S_s(f)$ symmetric w.r.t. the demodulation frequency f_0 , it is proved that the estimated DOAs given by any narrowband subspace-based algorithm are asymptotically unbiased w.r.t. the number of snapshots and signal bandwidth. More precisely

$$\mathbb{E}(\hat{\theta}_N) - \theta = \left(\frac{f_\sigma^2}{f_0^2} \right) \mathbf{b}^{\text{alg}} + O\left(\frac{f_\sigma^4}{f_0^4} \right) + O\left(\frac{1}{N} \right),$$

where $f_\sigma \stackrel{\text{def}}{=} [\int_{-B}^B S_s(f) f^2 df / \int_{-B}^B S_s(f) df]^{1/2}$ is the definition used for the bandwidth. Furthermore, for a single source, $\mathbf{R}_x = \mathbf{R}_{s_1} \odot \mathbf{a}(\theta_1) \mathbf{a}^H(\theta_1) + \sigma_n^2 \mathbf{I}$, where the nuisance parameters are now the terms of the Hermitian matrix \mathbf{R}_{s_1} and σ_n^2 . This new parameterization allows to derive the circular Gaussian stochastic CRB issued from a non-zero bandwidth $\text{CRB}_{\text{CG}}^{\text{NZB}}(\theta_1)$. It is related to the standard $\text{CRB}_{\text{CG}}(\theta_1)$ by the relation

$$\text{CRB}_{\text{CG}}^{\text{NZB}}(\theta_1) = \text{CRB}_{\text{CG}}(\theta_1) \left(1 + c \left(\frac{f_\sigma^2}{f_0^2} \right) + O\left(\frac{f_\sigma^4}{f_0^4} \right) \right),$$

where the expression of c is given in [20].

2) *Robustness to array modeling errors:* Imprecise knowledge of the gain and phase characteristics of the array sensors, and of the sensor locations and possible mutual coupling, can seriously degrade the theoretical performance of the DOA estimation algorithms. Experimental systems attempt to eliminate or minimize these errors by careful calibrations. But even when initial calibration is possible, system parameters may change over time and thus the array modeling errors cannot be completely eliminated. Consequently, it is useful to qualify the sensitivity of the DOA estimator algorithms to these modeling errors, i.e., to study the effect of difference between the true and assumed array manifold $\{\mathbf{a}(\theta), \theta \in \Theta\}$ caused by modeling errors, on DOA estimator algorithms. This analysis has received relatively little attention in the literature.

In these studies, to simplify the analysis, the covariance matrix \mathbf{R}_x is assumed perfectly known, i.e., the effects of a finite number of samples is assumed negligible. Let γ gather the array parameters which are the subject of the sensitivity analysis. For example, γ may contain the sensors gain, phases or location, or other parameters such as the mutual coupling coefficients of the array sensors. A DOA estimation algorithm uses the steering matrix $\mathbf{A}(\boldsymbol{\theta}, \gamma_0) = [\mathbf{a}(\theta_1, \gamma_0), \dots, \mathbf{a}(\theta_P, \gamma_0)]$, corresponding to a nominal value γ_0 of the array parameters that differs from the true steering matrix $\mathbf{A}(\boldsymbol{\theta}, \gamma)$, where γ is slightly different from γ_0 (see particular parameterizations studied in [36] and [30]). We refer to the difference between the true and assumed array parameters as a modeling error. The sensitivity study of a particular DOA estimation algorithm consists to provide a relation between $\delta\boldsymbol{\theta} = \boldsymbol{\theta}_\gamma - \boldsymbol{\theta}$ and the modeling error $\delta\gamma = \gamma - \gamma_0$ in the mapping

$$\mathbf{R}_x(\gamma) = \mathbf{A}(\boldsymbol{\theta}, \gamma)\mathbf{R}_s\mathbf{A}^H(\boldsymbol{\theta}, \gamma) + \sigma_n^2\mathbf{I} \xrightarrow{\text{alg}(\gamma_0)} \boldsymbol{\theta}_\gamma, \quad (71)$$

where naturally $\mathbf{R}_x(\gamma_0) \xrightarrow{\text{alg}(\gamma_0)} \boldsymbol{\theta}$, if $\text{alg}(\gamma_0)$ denotes an arbitrary second-order algorithm based on the nominal array. Using a first order perturbation of (71) in the neighborhood of γ_0 , through those of the orthogonal projector on the noise subspace $\mathbf{\Pi}_x(\gamma)$, a relation $\delta\boldsymbol{\theta} = h(\delta\gamma) + o(\delta\gamma)$ where h is linear has been given for the MUSIC and DML algorithms in [36], [90] and [37], respectively. These works model the errors $\delta\gamma$ by zero-mean independent random variables ($\delta\gamma = \sigma_\gamma\mathbf{u}$ where \mathbf{u} is a random vector whose elements are zero-mean unit variance random variables). They lead to estimates that are approximatively unbiased (i.e., $\mathbb{E}(\boldsymbol{\theta}_\gamma) - \boldsymbol{\theta} = o(\sigma_\gamma)$) and where their approximative variances depend only on the second-order statistics of the modeling errors (more precisely $\text{Var}(\theta_{p,\gamma}) = c_p\sigma_\gamma^2 + o(\sigma_\gamma^2)$, $p = 1, \dots, P$). However, by confronting these theoretical results with numerical experiments, one notices that the MUSIC and DML algorithms are biased in the presence of multiple sources and these theoretic and experimental variances do not agree with larger modeling errors. More precisely, these theoretical results are valid only up to the point where the probability of resolution is close to one (see [31]).

To take into account these larger modeling errors, a more accurate relation between $\delta\boldsymbol{\theta}$ and $\delta\gamma$, based on a second-order expansion of $\mathbf{\Pi}_x(\gamma)$ around γ_0 (provided by a recursive n th order expansion of $\delta\mathbf{\Pi}_x$ w.r.t. $\delta\mathbf{R}_x$ [51]) as been given in [31] and [32] for analyzing the sensitivity of the MUSIC and DML algorithms to larger modeling errors. Modeling the errors $\delta\gamma$ as previously, an approximation of the bias $\mathbb{E}(\boldsymbol{\theta}_\gamma) - \boldsymbol{\theta}$ that depends on the second-order statistics of the modeling errors, and of the variance that now depends on the fourth-order statistics of the modeling errors, are given. These refined closed-form expressions can predict the actual performance observed by numerical experiments for larger modeling errors, in particular in the threshold regions of the MUSIC and DML algorithms.

Note that the sensitivity of DOA estimators to modeling errors of the noise covariance matrix, that includes the presence of undetected weak signals, has also been studied in the literature (see e.g., [94]). Finally, note that the combined effects of random array modeling errors and finite samples have been analyzed for the class of so-called signal subspace fitting (SSF) algorithms in [95]. In addition to deriving the first-order asymptotic expressions for the covariance of the estimation error, an additional weighting matrix has been introduced in (61) that has been optimized for any particular random array modeling errors.

F. High-order algorithms

When the sources are non Gaussian distributed, they convey valuable statistical information in their moments of order greater than two (this is in particular true when considering communications signals). In these circumstances, it makes sense to consider DOA estimation techniques using this higher order information. Of particular interest are the algorithms based on higher order cumulants of the measurements $\{\mathbf{x}(t)\}_{t_1, \dots, t_N}$ due to their additivity property in the sums of independent components. Furthermore, these cumulants show the distinctive property of being in a certain sense, insensitive to additive Gaussian noise, making it possible to devise consistent DOA estimates without it being necessary to know, to model or to estimate the noise covariance \mathbf{R}_n . As generally, the distributions of the sources are even, their odd order moments are zero and thus to cope with these signals, only the even high-order cumulants of the measurements are used.

Computational considerations dictate using mainly fourth-order cumulants. To use these approaches, we consider the assumptions of Subsection II-B, in which we add that the sources $\{\mathbf{s}_p(t)\}_{p=1, \dots, P}$ have nonvanishing

fourth-order cumulants. Furthermore, we assume that their moments are finite up to the eighth-order, to study the statistical performance of these algorithms.

Of course, there are many more quadruples than pairs of indices, and consequently a very large number of cumulants $\text{Cum}(x_i(t), x_j^*(t), x_k^*(t), x_l(t))$, $i, j, k, l = 1, \dots, M$ for circular sources (and more, $\text{Cum}(x_i(t), x_j^*(t), x_k(t), x_l(t))$ and $\text{Cum}(x_i(t), x_j(t), x_k(t), x_l(t))$, $i, j, k, l = 1, \dots, M$ for noncircular sources) can be exploited despite their redundancies, to identify the DOA parameters with unknown noise covariance. For example, for circular signals, the maximum set of nonredundant cumulants is

$$\text{Cum}(x_i(t), x_j^*(t), x_k^*(t), x_l(t)) \text{ with } 1 \leq i \leq M, 1 \leq l \leq i, 1 \leq j \leq i \text{ and } 1 \leq k \leq j.$$

The asymptotically minimum variance (AMV) algorithm (see Subsection III-C) based on a subset of fourth-order cumulants that can identify the DOA parameters, is the nonlinear least square algorithm (40) in which \mathbf{g}_N gathers the involved cumulants. To implement this AMV algorithm, one has to decide which cumulants should be included in \mathbf{g}_N . The best estimate would be obtained when all nonredundant cumulants are selected. This, however, may require excessive computations if M is large. However it is sufficient to deal with a reduced set of cumulants, although there do not seem to be any simple guidelines in this matter [69]. In practice, a good tradeoff between computational complexity and accuracy is to devise suboptimal algorithms that require an overall computational effort similar to the second-order algorithms, while retaining a fourth-order cumulants subset, sufficient for DOA identification. Such algorithms have been proposed in the literature such as the diagonal slice (DS), the contracted quadricovariance (CQ) and the so called 4-MUSIC [69] algorithms. The first two algorithms are fourth-order subspace-based algorithms built on the following rank defective $M \times M$ matrices

$$\begin{aligned} (\mathbf{Q}_x^{\text{DS}})_{i,j} &= \text{Cum}(x_i(t), x_j^*(t), x_j^*(t), x_j(t)) \\ (\mathbf{Q}_x^{\text{CQ}})_{i,j} &= \sum_{m=1}^M \text{Cum}(x_i(t), x_j^*(t), x_m^*(t), x_m(t)). \end{aligned}$$

They require $P < M$ sources and their statistical performance has been analyzed in [11] with the general framework explained in Subsection III-A. In particular, it is has been proved that for a single source and a ULA in spatially uniform white noise, these two fourth-order algorithms have similar performance to the MUSIC algorithm, except for low SNR, for which the MUSIC algorithm outperforms both fourth-order algorithms. The 4-MUSIC algorithm is built from the rank defective $M^2 \times M^2$ matrix

$$(\mathbf{Q}_x^{4\text{-MUSIC}})_{i+(j-1)M, k+(l-1)M} = \text{Cum}(x_i(t), x_j^*(t), x_k^*(t), x_l(t)).$$

It is proved in [69] that

$$\mathbf{Q}_x^{4\text{-MUSIC}} = [\mathbf{A}^*(\boldsymbol{\theta}) \otimes \mathbf{A}(\boldsymbol{\theta})] \mathbf{Q}_s [\mathbf{A}^*(\boldsymbol{\theta}) \otimes \mathbf{A}(\boldsymbol{\theta})]^H,$$

where $(\mathbf{Q}_s)_{i+(j-1)P, k+(l-1)P} = \text{Cum}(s_i(t), s_j^*(t), s_k^*(t), s_l(t))$, $i, j, k, l = 1, \dots, P$. $\mathbf{Q}_x^{4\text{-MUSIC}}$ is indefinite in general and its rank is $\sum_{g=1}^G r_g^2$ where the P sources are divided in G groups, with r_g in the g th group. The sources in each group are assumed to be dependent, while sources belonging to different groups are assumed independent. Because the vectors $\mathbf{a}^*(\theta_p) \otimes \mathbf{a}(\theta_p)$, $p = 1, \dots, P$ are P columns of $\mathbf{A}^*(\boldsymbol{\theta}) \otimes \mathbf{A}(\boldsymbol{\theta})$, the 4-MUSIC algorithm is obtained by searching the P deepest minima of the following localisation function $J_{4\text{-MUSIC}}[\boldsymbol{\theta}, \mathbf{\Pi}_{x,N}]$

$$J_{4\text{-MUSIC}}[\boldsymbol{\theta}, \mathbf{\Pi}_{x,N}] = [\mathbf{a}^*(\boldsymbol{\theta}) \otimes \mathbf{a}(\boldsymbol{\theta})]^H \mathbf{\Pi}_{x,N} [\mathbf{a}^*(\boldsymbol{\theta}) \otimes \mathbf{a}(\boldsymbol{\theta})], \quad (72)$$

where $\mathbf{\Pi}_{x,N}$ is now, the orthogonal projector onto the noise subspace of the sample estimate $\mathbf{Q}_{x,N}^{4\text{-MUSIC}}$ of $\mathbf{Q}_x^{4\text{-MUSIC}}$. In practice the statistical dependence of the sources are unknown. [69] has proposed to retain only $M^2 - P^2$, rather $M^2 - \sum_{g=1}^G r_g^2$ eigenvectors corresponding to the smallest singular values of $\mathbf{Q}_{x,N}^{4\text{-MUSIC}}$. We note that, to the best of our knowledge, no complete statistical performance analysis of this algorithm has yet appeared in the literature. Despite its higher variance (w.r.t. the MUSIC algorithm under the assumption of spatially uniform white noise), this fourth-order algorithm presents some advantages, aside from its capacity to deal with unknown Gaussian noise fields. Using the concept of virtual array, it is proved in [13] that this

algorithm can identify up to $M^2 - M$ sources when the sensors are identical and up to $M^2 - 1$ sources for different sensors. Furthermore, it is shown that its resolution for closely spaced sources and robustness to modeling errors is improved with respect to the MUSIC algorithm. To increase even more its number of sources to be processed, resolution and robustness to modeling errors, extensions of this 4-MUSIC algorithms, giving rise to the $2q$ -MUSIC (with $q > 2$) has been proposed [14].

V. DETECTION OF NUMBER OF SOURCES

One of the more difficult and critical problems facing passive sensor arrays systems is the detection of the number P of sources impinging on the array. This is a key step in most of the parametric estimation techniques that were briefly described in Section IV. The eigendecomposition based techniques require in addition, information on the dimension r of the signal subspace. If the source covariance \mathbf{R}_s has full rank, i.e., there are no coherent sources present, P and r are identical. Moreover, the solution of the detection problem has, in many cases, value of its own, regardless of the DOA estimation problem.

A natural scheme for detecting the number P of sources is to formulate a likelihood ratio test based on the SML estimator (49). Such a test is often referred to as a generalized likelihood ratio test (GLRT). This test can be implemented by a sequential test procedure (see e.g., [64, Sec. 4.7.1]). For each hypothesis, the likelihood ratio statistic is computed and compared to a threshold. The accepted hypothesis is the first one for which the threshold is crossed. The problem with this method is the subjective judgment required for deciding on the threshold levels or the associated probabilities of false alarm related by the asymptotic distribution of the normalized likelihood ratio.

Another important approach to the detection problem is the application of the information theoretic criteria for model selection. Unlike the conventional hypothesis testing based approaches, these criteria do not require any subjective threshold setting. Among them, the minimum description length (MDL) criterion introduced by Schwartz and Rissanen [73] is the most widely used because of its consistency. This technique has been used for detecting the signal subspace dimension r [99], and also for detecting the number of sources P [101]. We concentrate now on the detection of r .

A. MDL criterion

The information theoretic criteria approach is a general method for detecting the order r of a statistical model. That is, given a parameterized probability density function $p(\mathbf{x}; \boldsymbol{\alpha}^{(r)})$ for various order r , detect \hat{r} such that $\hat{r} = \arg \min_r \{-\ln[p(\mathbf{x}; \hat{\boldsymbol{\alpha}}_{\text{ML}}^{(r)})] + g(r)\}$, where $\hat{\boldsymbol{\alpha}}_{\text{ML}}^{(r)}$ is the ML estimate of $\boldsymbol{\alpha}^{(r)}$ and $g(r)$ is a penalty function. For the MDL criterion which is based on a particular penalty function, \hat{r} is given for N independent identically distributed measurements $\mathbf{x}(t_n)$, by

$$\hat{r} = \arg \min_r \left[-\ln \left(p(\mathbf{x}; \hat{\boldsymbol{\alpha}}_{\text{ML}}^{(r)}) \right) + \frac{1}{2} \text{card}(\boldsymbol{\alpha}^{(r)}) \ln(N) \right], \quad (73)$$

where $\text{card}(\boldsymbol{\alpha}^{(r)})$ denotes the number of free real-valued parameters in $\boldsymbol{\alpha}^{(r)}$. Depending on the distribution of the measurements \mathbf{x} and its parametrization $\boldsymbol{\alpha}$, different implementations of the MDL criterion have been proposed.

The most often used assumption, is the zero-mean circular Gaussian distribution associated with the parametrization (1) in which all the elements of the steering matrix \mathbf{A} are assumed unknown with the only restriction that \mathbf{A} has full column rank with $M > P$. For this modeling, the measurements can be parameterized by the parameter

$$\boldsymbol{\alpha}^{(r)} = [\mathbf{v}_1^T, \dots, \mathbf{v}_r^T, \lambda_1, \dots, \lambda_r, \sigma_n^2]^T,$$

where $\lambda_1 \geq \dots \geq \lambda_r > \sigma_n^2 = \dots, \sigma_n^2$ are the eigenvalues of \mathbf{R}_x and $\mathbf{v}_1, \dots, \mathbf{v}_r$, the eigenvectors associated with the largest r eigenvalues, and the general MDL criterion (73), which is referred to as the Gaussian MDL (GMDL), becomes [99]

$$\hat{r} = \text{Arg} \min_r \Lambda_r \quad \text{with} \quad \Lambda_r \stackrel{\text{def}}{=} N(M-r) \ln \left(\frac{\hat{a}_r}{\hat{g}_r} \right) + \frac{1}{2} r(2M-r) \ln N, \quad (74)$$

with $\hat{a}_r \stackrel{\text{def}}{=} \frac{1}{M-r} \sum_{i=r+1}^M \hat{\lambda}_i$ and $\hat{g}_r \stackrel{\text{def}}{=} \prod_{i=r+1}^M \hat{\lambda}_i^{1/(M-r)}$, where $\hat{\lambda}_1 > \hat{\lambda}_2 > \dots > \hat{\lambda}_M$ are the eigenvalues of the sample covariance matrix $\frac{1}{N} \sum_{n=1}^N \mathbf{x}(t_n) \mathbf{x}^H(t_n)$, denoted here by $\hat{\mathbf{R}}_x$.

B. Performance analysis of MDL criterion

This GMDL criterion has been analyzed in [106], and it has been shown to be a consistent estimator of the rank r , i.e., the probability of error decreases to zero as the number N of measurements increases to infinity. Moreover, under mild regularity conditions, like finite second moments, it is a consistent estimator of the rank r , even if the measurements are non-Gaussian. This property contrasts with the Akaike information criterion (AIC) that yields an inconsistent estimate of that tends, asymptotically, to overestimate r [99].

The GMDL criterion has been further analyzed by considering the events $\hat{r} < r$ and $\hat{r} > r$, called underestimation and overestimation, respectively. Since $(\Lambda_r)_{r=0, \dots, M-1}$ are functions of the eigenvalues $(\hat{\lambda}_i)_{i=1, \dots, M}$ of $\hat{\mathbf{R}}_x$, the derivation of the probabilities $P(\hat{r} > r)$ and $P(\hat{r} < r)$ needs the joint exact or asymptotic distribution of $(\hat{\lambda}_i)_{i=1, \dots, M}$. This asymptotic distribution is available for circular complex Gaussian distribution [6] and more generally for arbitrary distributions with finite fourth-order moments [26], but unfortunately, the functional $(\Lambda_r)_{r=0, \dots, M-1}$ (74) is too complicated to infer its asymptotically distribution. Therefore, for simplifying the derivation of these probabilities, it has been argued [98], [48] and [104] by extended Monte Carlo experiments (essentially for $r = 1$ and $r = 2$) that

$$P(\hat{r} > r) \approx P(\hat{r} = r + 1) \approx P(\Lambda_{r+1} < \Lambda_r) \quad \text{and} \quad P(\hat{r} < r) \approx P(\hat{r} = r - 1) \approx P(\Lambda_{r-1} < \Lambda_r). \quad (75)$$

As the probability of overestimation is concerned, exact and approximate asymptotic upper bound of this probability have been derived in [104] showing that generally $P(\hat{r} > r) \ll 1$. Therefore, only the probability of underestimation has been analyzed by many authors. In particular, using the refinement introduced by [41]

$$E(\hat{a}_r) = \frac{1}{M-r} (\text{Tr}(\mathbf{R}_x) - \sum_{i=1}^r E(\hat{\lambda}_i)) = \sigma_n^2 + \frac{1}{M-r} \sum_{i=1}^r (\lambda_i - E(\hat{\lambda}_i))$$

of the classical approximation $E(\hat{a}_r) \approx \sigma_n^2$ and the asymptotic bias (18) and covariance (19), a closed-form expression of the probability of underestimation given by the GMDL criterion, used under arbitrary distributions with finite fourth-order moments, has been given in [26]. This expression has been analyzed for $P = r = 1$ and $P = r = 2$ for different distributions of the sources in [26]. Fig.5 illustrates the robustness of the MDL criterion to the distribution of the source. We see from this figure that the probability of underestimation is sensitive to the distribution of the source, particularly for sources of large kurtosis and for weak values of the number N of snapshots.

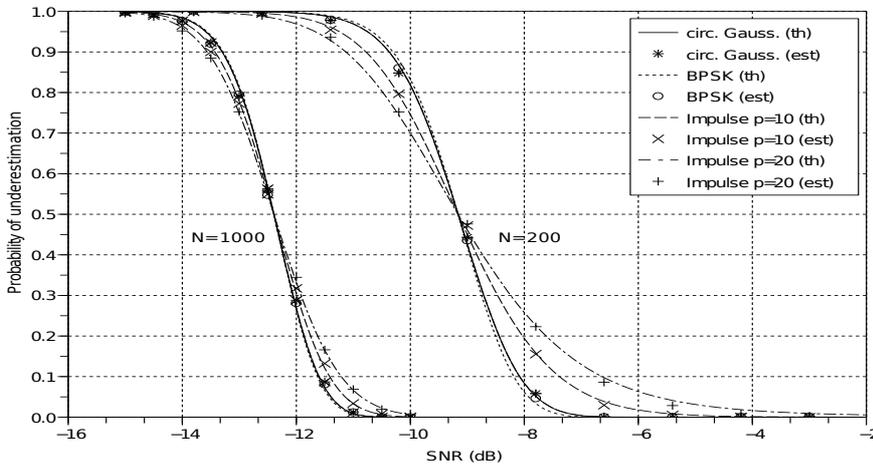


Fig.5 $P(\hat{r} = 0/r = 1)$ as a function of the SNR for four distributions of the source (the impulsive takes the values $\{-1, 0, +1\}$ with $P(s(t_n) = -1) = P(s(t_n) = +1) = \frac{1}{2p}$) and two values of the number N of snapshots, for an ULA with 5 sensors.

The general MDL criterion has been studied in [33], where using the approximation (75), a general analytical expression of $P(\hat{r} < r)$ has been given. This expression allows one to prove the consistency of the general MDL criterion when the number of snapshots tends to infinite and has been specialized to particular parameterized distributions. Among them, the Gaussian assumption associated with a parameterized steering matrix $\mathbf{A}(\boldsymbol{\theta})$ has been studied and some numerical illustrations show that the use of this prior information about the array geometry enables an improvement in performance of about 2dB. Finally, note that the MDL criterion generally fails when the sample size is smaller than the number of sensors. In this situation a sample eigenvalue based detector has been proposed in [60].

VI. RESOLUTION OF TWO CLOSELY SPACED SOURCES

An important measure to quantify the statistical performance for the DOA estimation problem is the resolvability of closely spaced signals in terms of their parameters of interest. The principal question to characterize this resolvability is to find the minimum SNR (denoted threshold array SNR (ASNR)) required for a sensor array to correctly resolve two closely spaced signals for a given DOA distance $\Delta\theta \stackrel{\text{def}}{=} |\theta_2 - \theta_1|$ (called angular resolution limit (ARL) or statistical resolution limit) between them. Generally in the literature they are three different ways to describe this resolution limit. The first one is based on the mean null spectrum concerning a specific algorithm. the second one is based on the estimation accuracy, more precisely on the Cramer-Rao Bound. The last one is based on the detection theory using the hypothesis test formulation.

A. Angular resolution limit based on mean null spectra

Based on the array beam-pattern $G(\theta_0, \theta) = |\mathbf{a}^H(\theta_0)\mathbf{a}(\theta)|$, different resolution criteria have been defined from its main lobe w.r.t. a look direction θ_0 , as the celebrated Rayleigh resolutions such as the half power beamwidth or the null to null beamwidth that depends solely on the antenna geometry, and consequently have the serious shortcoming of being independent of the SNR.

For specific so-called high resolution algorithms, such as different MUSIC-like algorithms, based on the search for two local minima of sample null spectra $J_{\text{Alg}}(\theta, \mathbf{\Pi}_{x,N})$, two main criteria based on the mean null spectrum $\mathbb{E}[J_{\text{Alg}}(\theta, \mathbf{\Pi}_{x,N})]$ have been defined. These criteria are justified by the property that the standard deviation $\sqrt{\text{Var}(J_{\text{Alg}}(\theta, \mathbf{\Pi}_{x,N}))}$ of the sample null spectrum associated with the conventional MUSIC and Min-Norm algorithms is small compared to its mean value $\mathbb{E}[J_{\text{Alg}}(\theta, \mathbf{\Pi}_{x,N})]$ in the vicinity of the true DOAs for $N \gg M$ for arbitrary SNR [47].

For the first criterion, introduced by Cox [15], two sources are resolved if the midpoint mean null spectrum is greater than the mean null spectrum in the two true source DOAs.

$$\mathbb{E}[J_{\text{Alg}}(\theta_m, \mathbf{\Pi}_{x,N})] \geq \frac{1}{2} (\mathbb{E}[J_{\text{Alg}}(\theta_1, \mathbf{\Pi}_{x,N})] + \mathbb{E}[J_{\text{Alg}}(\theta_2, \mathbf{\Pi}_{x,N})]) \quad \text{with } \theta_m \stackrel{\text{def}}{=} \frac{1}{2}(\theta_1 + \theta_2).$$

This criterion was first studied by Kaveh and Barabell [47], [49] in the resolution analysis of the conventional MUSIC and Min-Norm algorithms for two uncorrelated equal-powered sources and a ULA. This analysis has been extended to more general classes of situations, e.g., for two correlated or coherent equal-powered sources with the smoothed MUSIC algorithm [66], then for two unequal-powered sources impinging on an arbitrary array with the conventional and beamspace MUSIC algorithm [53]. A subsequent paper by Zhou *et al.* [109] developed a resolution measure based on the mean null spectrum and compared their results to Kaveh and Barabell's work.

For the second criterion, introduced by Sharman and Durrani [76] and then studied by Forster and Villier [34] in the context of the conventional MUSIC and Min-Norm algorithms for two uncorrelated equal-powered sources and a ULA, two sources are resolved if the second derivative of the mean null spectrum at the midpoint is negative.

$$\frac{d^2 \mathbb{E}[J_{\text{Alg}}(\theta, \mathbf{\Pi}_{x,N})]}{d\theta^2} \Big|_{\theta=\theta_m} \leq 0.$$

Resorting to an analysis based on perturbations of the noise projector $\mathbf{\Pi}_{x,N}$ [51], instead of those of the eigenvectors (e.g., [47] and [53]), these two criteria have been studied for arbitrary distributions of the sources,

for the conventional MUSIC algorithm. The following closed-form expressions of the approximation of the threshold ASNR given by these two criteria have been obtained in [3]:

$$\text{ASNR}_1 \approx \frac{2}{N} \frac{\alpha_M}{(\Delta\theta)^4} \left(1 + \sqrt{1 + \frac{N(\Delta\theta)^2}{2\beta_M}} \right) \quad \text{and} \quad \text{ASNR}_2 \approx \frac{1}{N} \frac{\alpha_M}{(\Delta\theta)^4} \left(1 + \sqrt{1 + \frac{N(\Delta\theta)^2}{\beta_M}} \right), \quad (76)$$

where α_M and β_M are fractional expressions in M specified in [3] for ULAs. These expressions (76) have been extended in [3] to a noncircular MUSIC algorithm adapted to rectilinear signals, introduced and analyzed in [5], for which (76) becomes

$$\text{ASNR}_1 \approx \frac{2}{N} \alpha_M^{\Delta\theta, \Delta\phi} \left(1 + \sqrt{1 + \frac{N}{2\beta_M^{\Delta\theta, \Delta\phi}}} \right) \quad \text{and} \quad \text{ASNR}_2 \approx \frac{1}{N} \alpha_M^{\Delta\theta, \Delta\phi} \left(1 + \sqrt{1 + \frac{N}{\beta_M^{\Delta\theta, \Delta\phi}}} \right), \quad (77)$$

where $\Delta\phi \stackrel{\text{def}}{=} \phi_2 - \phi_1$ is the second-order noncircularity phase separation (8) and where now $\alpha_M^{\Delta\theta, \Delta\phi}$ and $\beta_M^{\Delta\theta, \Delta\phi}$ are expansions of $1/(\Delta\theta)^2$ without constant term, whose coefficients depend on M , $\Delta\phi$ and the array configuration. Closed-form expressions of $\alpha_M^{\Delta\theta, \Delta\phi}$ and $\beta_M^{\Delta\theta, \Delta\phi}$ are given in [3] for weak and large second-order noncircularity phase separations and ULAs, where it is proved that ASNR_1 and ASNR_2 are decreasing functions of $\Delta\phi$ and thus are minimum for $\Delta\phi = \pi/2$.

Fig.6 illustrates these two threshold ASNRs for two independent equal-powered BPSK modulated signals impinging on a ULA with $M = 10$ and $T = 500$. We clearly see in this figure that the noncircular MUSIC algorithm outperforms the conventional MUSIC algorithm except for very weak second-order noncircularity phase separations for which the ASNR thresholds of these two algorithms are very similar. Furthermore, we note that the behaviors of the ASNR threshold given by the two criteria are very similar although the ASNR thresholds are slightly weaker for the Sharman and Durrani criterion than for the Cox criterion.

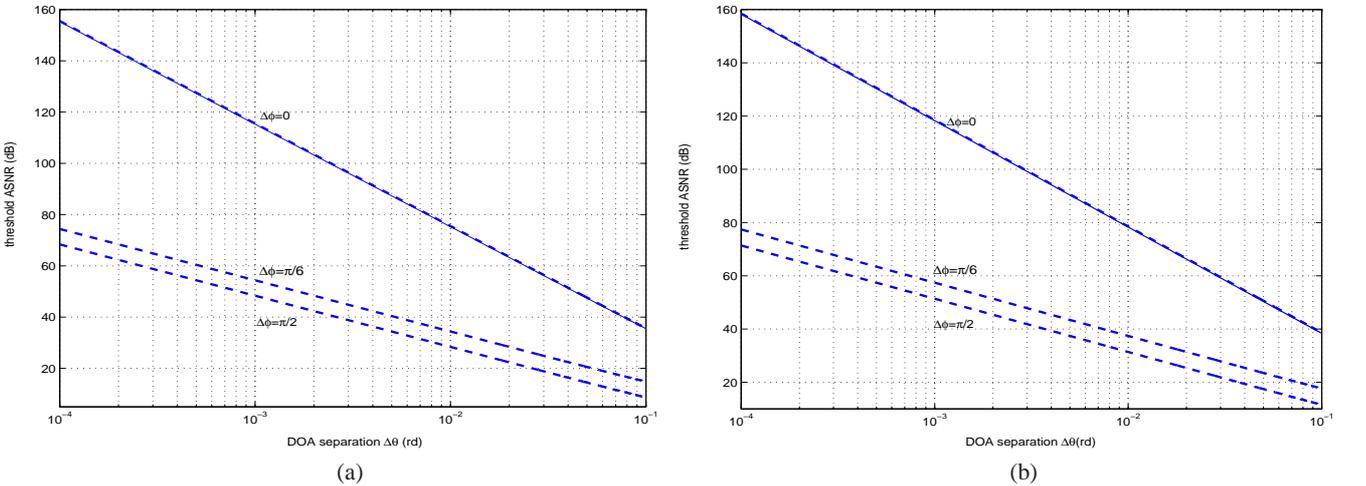


Fig.6 Comparison of the threshold ASNRs given by the Cox (a) and Sharman and Durrani (b) criteria as a function of the DOA separation $\Delta\theta$ associated with the conventional MUSIC (—) and noncircular MUSIC algorithms (- -) for three values of the second-order noncircularity phase separation $\Delta\phi$.

Moreover, several authors have considered (e.g., [54], [107], [108]) the probability of resolution or an approximation of it, based on the Cox criterion applied to the null sample spectrum to circumvent the possible misleading results given by these two criteria. Finally note that the resolution capability of the conventional and Capon beamforming algorithms have been thoroughly analyzed (see e.g., [72]). Thanks to the simple expression of their spatial null spectra (46), it is possible to derive an approximation of the probability of resolution defined as the probability that the dip in midway between the two sources is at least 3-dB less than the peak of either source as a function of the SNR and DOA separation. Thus, fixing a specific high confidence level, this allows one to predict the SNR required to resolve two closely spaced sources. The superiority of the Capon algorithm is proved in [72], as the resolving power increases with SNR; in contrast, the Bartlett algorithm cannot exceed

the Fourier/Rayleigh limit no matter how strong the signals.

B. Angular resolution limit based on the CRB

Array resolution has been studied independently of any algorithm by using the CRB. Based on the observation that the standard MUSIC algorithm is unlikely to resolve closely spaced signals if the standard deviation of the DOA estimates exceed $\Delta\theta/8$ [80], Lee [55] has proposed to define the resolution limit as the DOA separation $\Delta\theta$ for which

$$\max\left(\sqrt{\text{CRB}(\theta_1)}, \sqrt{\text{CRB}(\theta_2)}\right) = c\Delta\theta, \quad (78)$$

for the two closely spaced sources, where c is somewhat arbitrarily chosen. This criterion ignores the coupling between the estimates $\hat{\theta}_1$ and $\hat{\theta}_2$. To overcome these drawbacks, Smith has proposed [78] to define the resolution limit as the source separation that equals the square root of its own CRB, i.e.,

$$\sqrt{\text{CRB}(\Delta\theta)} = c\Delta\theta, \quad (79)$$

with $c = 1^9$. This means that the angular resolution limit or the threshold ASNR are obtained by resolving the implicit equations (78) and (79). This latter criterion has been applied to the deterministic modeling of the sources in [78] and then extended to multiple parameters per source in [29]. For the stochastic modeling of the sources, the circular Gaussian distribution has been compared to the discrete one in [24]. In particular it has been proved that the threshold ASNR is inversely proportional to the number N of snapshots and to the square of $\Delta\theta$ for the Gaussian case, in contrast to BPSK, MSK and QPSK case, for which it is inversely proportional to the fourth power of $\Delta\theta$.

C. Angular resolution limit based on the detection theory

The previous two approaches to characterize the angular resolution have in fact two different purposes. The first one studies the capability of a specific algorithm to estimate the DOAs of two closely spaced sources when the number of sources is known. In contrast, the second one is aiming to define an absolute limit on resolution that depends only of the array configuration and parameters of interest as the number M of sensors and SNR. But this latter approach based on the ad-hoc relationships (78) and (79), essentially makes sense because the CRB indicates the parameter estimation accuracy and intuitively should be related to the resolution limit. But it suffers from two drawbacks. First, the resolution limit defined by this approach is not rigorously grounded in a statistical setting. Second, if the resolution limit is expressible by (78) or (79), can the translation factor c , be analytically determined?

To solve these two problems, Liu and Nehorai have proposed to use a hypothesis test formulation [59]. This approach has been introduced in a 3D reference frame, but to be consistent with the notations of this section, it is briefly summarized in the following in the 2D framework, where the DOA of a source is the parameter θ . As the source localization accuracy may vary at different DOAs, consider the resolution limit at a specific DOA of interest. More precisely, assume there exists a source at a known DOA θ_1 and we are interested in the minimum angular separation $\Delta\theta$ that the array can resolve between this source at θ_1 and another source at a direction θ_2 close to θ_1 . Quite naturally, the resolution of the two sources can be achieved through the binary composite hypothesis test

$$\begin{aligned} H_0 & : \Delta\theta = 0, & \text{one source is present} \\ H_1 & : \Delta\theta > 0, & \text{two sources are present.} \end{aligned}$$

To rigorously define the resolution limit $\Delta\theta$, we fix the values of P_{FA} and P_{D} for this test. Otherwise, $\Delta\theta$ could be arbitrary low, while the result of the test may be meaningless. Let $\alpha = [\Delta\theta, \beta^T]^T$ be the unknown parameter of our statistical model, where $\Delta\theta$ is the parameter of interest and β gathers all the unknown nuisance

⁹Note that this translation factor c is somewhat arbitrarily chosen (see different values cited in [59]).

parameters. To conduct this test, the GLRT is considered due to the unknown nuisance parameters.

$$L_G(\mathbf{x}, N) = \frac{p(\mathbf{x}; \widehat{\Delta\theta}, \widehat{\boldsymbol{\beta}}_1, H_1)}{p(\mathbf{x}; \widehat{\boldsymbol{\beta}}_0, H_0)} \stackrel{H_1}{>} \gamma', \quad (80)$$

where $p(\mathbf{x}; \Delta\theta, \boldsymbol{\beta}, H_1)$ and $p(\mathbf{x}; \boldsymbol{\beta}, H_0)$ denote the probability density function of the measurement $\mathbf{x} = [\mathbf{x}^T(t_1), \dots, \mathbf{x}^T(t_N)]^T$ under the hypothesis H_1 and H_0 , respectively. $\widehat{\Delta\theta}$ and $\widehat{\boldsymbol{\beta}}_1$ are respectively the ML estimate of $\Delta\theta$ and $\boldsymbol{\beta}$ under H_1 , and $\widehat{\boldsymbol{\beta}}_0$ is the ML estimate of $\boldsymbol{\beta}$ under H_0 . The distribution of this GLRT $L_G(\mathbf{x}, N)$ is generally very involved to derive, but hopefully, approximations of the distribution of $2 \ln L_G(\mathbf{x}, N)$ for large values of N are available under H_0 and H_1 . First, under H_0 , Wilk's theorem with nuisance parameters (see e.g., [102, p.132]) can be applied without having to know the exact form of $L_G(\mathbf{x}, N)$. This theorem states the following convergence in distribution when N tends to ∞

$$2 \ln L_G(\mathbf{x}, N) \stackrel{L}{\rightarrow} \chi^2(1) \quad \text{under } H_0, \quad (81)$$

where $\chi^2(1)$ denotes the central chi-square distribution with one degree of freedom (associated with the single parameter $\Delta\theta$). Under H_1 , the derivation of the asymptotic distribution of $2 \ln L_G(\mathbf{x}, N)$ is much more involved. Using a theoretical result by Stroud [88], Stuart *et al* [89, Ch. 23.7] have stated that when $\Delta\theta$ can take values¹⁰ near $\mathbf{0}$, $2 \ln L_G(\mathbf{x}, N)$ is approximately distributed¹¹ as

$$2 \ln L_G(\mathbf{x}, N) \stackrel{a}{\sim} \chi^2(1, \lambda_N) \quad \text{under } H_1, \quad (82)$$

where $\chi^2(1, \lambda_N)$ denotes the noncentral chi-squared distribution with 1 degree of freedom and noncentrality parameter λ_N given by (see [50, Section 6.5])

$$\lambda_N = (\Delta\theta - 0)([\mathbf{FIM}^{-1}(\boldsymbol{\alpha})]_{1,1})^{-1}(\Delta\theta - 0), \quad (83)$$

whose dependence on N in the FIM of $\boldsymbol{\alpha}$ is emphasized, and where $[\mathbf{FIM}^{-1}(\boldsymbol{\alpha})]_{1,1}$ denotes the (1,1)th entry of $\mathbf{FIM}^{-1}(\boldsymbol{\alpha})$. It is further shown ([50, App. 6C]) that as N is large, (83) is approximated by

$$\lambda_N \approx (\Delta\theta)^2([\mathbf{FIM}^{-1}(\boldsymbol{\alpha})]_{1,1}|_{\Delta\theta=0})^{-1} = \text{CRB}^{-1}(\Delta\theta)|_{\Delta\theta=0}. \quad (84)$$

Based on these limit and approximate distributions of $2 \ln L_G(\mathbf{x}, N)$ under H_0 and H_1 for which the GLRT in (81) can be rewritten as

$$2 \ln L_G(\mathbf{x}, N) \stackrel{H_1}{>} \gamma \stackrel{\text{def}}{=} 2 \ln \gamma', \quad (85)$$

the angular resolution limit (ARL) has been computed in [59] by using the two constraints

$$P_{\text{FA}} = Q_{\chi^2(1)}(\gamma) \quad \text{and} \quad P_{\text{D}} = Q_{\chi^2(1, \lambda_N)}(\gamma),$$

where the values of P_{FA} and P_{D} are fixed and where $Q_{\chi^2(1)}$ and $Q_{\chi^2(1, \lambda_N)}$ denote the right tail probability of the $\chi^2(1)$ and $\chi^2(1, \lambda_N)$ distributions, respectively. It assumes the form

$$\Delta\theta = \sqrt{\lambda_K} \sqrt{\text{CRB}(\Delta\theta)|_{\Delta\theta=0}},$$

where the factor $\sqrt{\lambda_K}$ is analytically determined by the preassigned values of P_{FA} and P_{D} . Note that the SNR is embedded in the expression of $\text{CRB}(\Delta\theta)$ that is proportional to K . The dependence on the SNR of the CRB may vary according to the distribution of the sources. For example, [24] proves that the CRB of the DOA separation of discrete sources is very different from those of Gaussian sources.

¹⁰The following more formal condition is given in [88], $\Delta\theta$ is embedded in an adequate sequence indexed by N that converges to zero at the rate $N^{-1/2}$ or faster, i.e., $\|\Delta\theta\| = O(1/N^{1/2})$. Note the simplified condition given by Kay [50, A. 6A]: $\|\Delta\theta\| = c/\sqrt{N}$ for some constant c , that is reduced to the rough assumption of weak SNR [50, Section 6.5].

¹¹The accurate formulation is $\lim_{N \rightarrow \infty} \{P(2 \ln L_G(\mathbf{x}, N) < t) - P(V_N < t)\} = 0 \quad \forall t$, where V_N has a noncentral chi-squared distribution with one degree of freedom and noncentrality parameter μ_N that depends on the data length N .

REFERENCES

- [1] H. Abeida and J.P. Delmas, "Asymptotically minimum variance estimator in the singular case," *Proc. EUSIPCO, Antalya*, Sept. 2005.
- [2] H. Abeida and J.P. Delmas, "Cramer-Rao bound for direction estimation of non-circular signals in unknown noise fields," *IEEE Trans. Signal Process.*, vol. 53, no. 12, pp. 4610-4618, Dec. 2005.
- [3] H. Abeida and J.P. Delmas, "MUSIC-like estimation of direction of arrival for non-circular sources," *IEEE Trans. Signal Process.*, vol. 54, no. 7, pp. 2678-2690, July 2006.
- [4] H. Abeida and J.P. Delmas, "Efficiency of subspace-based DOA estimators," *Signal Processing (Elsevier)*, vol. 87, no. 9, pp. 2075-2084, Sep. 2007.
- [5] H. Abeida and J.P. Delmas, "Statistical performance of MUSIC-like algorithms in resolving noncircular sources," *IEEE Trans. Signal Process.*, vol. 56, no. 9, pp. 4317-4329, Sep. 2008.
- [6] T.W. Anderson, "Asymptotic theory for principal component analysis," *Ann. Math. Statist.*, vol. 34, pp. 122-148, 1963.
- [7] W.J. Bangs "Array processing with generalized beamformers," *Ph.D. thesis Yale University, New Haven, CT*, 1971.
- [8] Y. Begriche, M. Thameri, and K. Abed-Meraim, "Exact Cramer Rao bound for near field source localization," *International conference on ISSPA*, Montreal, July 2012.
- [9] F. Bellili, S.B. Hassen, S. Affes, and A. Stephenne, "Cramer-Rao Lower Bounds of DOA Estimates from Square QAM-Modulated Signals," *IEEE Trans. Commun.*, vol. 59, no. 6, pp. 1675-1685, June 2011.
- [10] J.F. Cardoso and E. Moulines, "A robustness property of DOA estimators based on covariance," *IEEE Trans. Signal Process.*, vol. 42, no. 11, pp. 3285-3287, Nov. 1994.
- [11] J.F. Cardoso and E. Moulines, "Asymptotic performance analysis of direction-finding algorithms based on fourth-order cumulants," *IEEE Trans. Signal Process.*, vol. 43, no. 1, pp. 214-224, Jan. 1995.
- [12] J.F. Cardoso and E. Moulines, "Invariance of subspace based estimator," *IEEE Trans. Signal Process.*, vol. 48, no. 9, pp. 2495-2505, Sep. 2000.
- [13] P. Chevalier and A. Ferreol, "On the virtual array concept for the fourth-order direction finding problem," *IEEE Trans. Signal Process.*, vol. 47, no. 9, pp. 2592-2595, Sept. 1999.
- [14] P. Chevalier, A. Ferreol, and L. Albera, "High resolution direction finding from higher order statistics; the 2q-MUSIC algorithm," *IEEE Trans. Signal Process.*, vol. 54, no. 8, pp. 2986-2997, August 2006.
- [15] H. Cox, "Resolving power and sensitivity to mismatch of optimum array processors," *J. Acoust. Soc. America*, vol. 54, no. 3, pp. 771-785, 1973.
- [16] J.P. Delmas, J.F. Cardoso, "Performance analysis of an adaptive algorithm for tracking dominant subspace," *IEEE Transactions on Signal Processing*, vol. 46, no. 11, pp. 3045-3057, November 1998.
- [17] J.P. Delmas, J.F. Cardoso, "Asymptotic distributions associated to Oja's learning equation for Neural Networks," *IEEE Transactions on Neural Networks*, vol. 9, no. 6, pp. 1246-1257, November 1998.
- [18] J.P. Delmas and Y. Meurisse, "Asymptotic performance analysis of DOA algorithms with temporally correlated narrow-band signals," *IEEE Trans. Signal Process.*, vol. 48, no. 9, pp. 2669-2674, Sep. 2000.
- [19] J.P. Delmas, "Asymptotic performance of second-order algorithms," *IEEE Trans. Signal Process.*, vol. 50, no. 1, pp. 49-57, Jan. 2002.
- [20] J.P. Delmas and Y. Meurisse, "Robustness of narrowband DOA algorithms with respect to signal bandwidth," *Signal Processing*, vol. 83, no. 3, pp. 493-510, Feb. 2003.
- [21] J.P. Delmas, "Asymptotically minimum variance second-order estimation for non-circular signals with application to DOA estimation," *IEEE Trans. Signal Processing*, vol. 52, no. 5, pp. 1235-1241, May 2004.
- [22] J.P. Delmas and H. Abeida, "Stochastic Cramer-Rao bound for non-circular signals with application to DOA estimation," *IEEE Trans. Signal Process.*, vol. 52, no. 11, pp. 3192-3199, Nov. 2004.
- [23] J.P. Delmas and H. Abeida, "Cramer-Rao bounds of DOA estimates for BPSK and QPSK modulated signals," *IEEE Trans. Signal Process.*, vol. 54, no. 1, pp. 117-126, Jan. 2006.
- [24] J.P. Delmas, H. Abeida, "Statistical resolution limits of DOA for discrete sources," in *Proc. ICASSP*, Toulouse, May 2006.
- [25] J.P. Delmas and H. Abeida, "Asymptotic distribution of circularity coefficients estimate of complex random variables," *Signal Processing*, vol. 89, pp. 2670-2675, May 2009.
- [26] J.P. Delmas and Y. Meurisse, "On the second-order statistics of the EVD of sample covariance matrices - application to the detection of noncircular or/and nonGaussian components," *IEEE Trans. Signal Process*, vol. 59, no. 8, pp. 4017-4023, August 2011.
- [27] J.P. Delmas and H. Gazzah, "Analysis of near-field source localization using uniform circular arrays," in *International Conference on Acoustics, Speech and Signal Processing (ICASSP 2013)*, Vancouver, Canada, May 2013.
- [28] M.N. El Korso, R. Boyer, A. Renaux, and S. Marcos, "Conditional and Unconditional Cramer-Rao Bounds for Near-Field Source Localization," *IEEE Trans. Signal Process*, vol. 58, no. 5, pp. 2901-2907, May 2010.

- [29] M.N. El Korso, R. Boyer, A. Renaux, and S. Marcos, "Statistical resolution limit for multiple parameters of interest and for multiple signals," in *Proc. ICASSP*, Dallas, May 2010.
- [30] A. Ferreol, P. Larzabal, and M. Viberg, "On the asymptotic performance analysis of subspace DOA estimation in the presence of modeling errors: Case of MUSIC," *IEEE Trans. Signal Process.*, vol. 54, no. 3, pp. 907-920, March 2006.
- [31] A. Ferreol, P. Larzabal, and M. Viberg, "On the resolution probability of MUSIC in presence of modeling errors," *IEEE Trans. Signal Process.*, vol. 56, no. 5, pp 1945-1953, May 2008.
- [32] A. Ferreol, P. Larzabal, and M. Viberg, "Performance prediction of maximum likelihood direction of arrival estimation in the presence of modeling error," *IEEE Trans. Signal Process.*, vol. 56, no. 10, pp 4785-4793, Oct. 2008.
- [33] E. Fishler, M. Grossmann, and H. Messer, "Detection of signals by information theoretic criteria: general asymptotic performance analysis," *IEEE Trans. Signal Process.*, vol. 50, no. 5, pp. 1027-1036, May 2002.
- [34] P. Forster and E. Villier, "Simplified formulas for performance analysis of MUSIC and Min Norm," *Proceedings of Ocean conf.*, Sep. 1998.
- [35] B. Friedlander and B. Porat, "Asymptotically optimal estimation of MA and ARMA parameters of non-Gaussian processes from high-order moments," *IEEE Trans. Automat. Contr.*, vol. 35, pp. 27-35, Jan. 1990.
- [36] B. Friedlander, "A sensitivity analysis of the MUSIC algorithm," *IEEE Trans. on ASSP*, vol. 38, no. 10, pp. 1740-1751, Oct. 1990.
- [37] B. Friedlander, "A sensitivity analysis of the maximum likelihood direction-finding algorithm," *IEEE Trans. Aerosp. Electro. Syst.*, vol. 26, no. 11, pp. 953-958, Nov. 1990.
- [38] H. Gazzah and J.P. Delmas, "Spectral efficiency of beamforming-based parameter estimation in the single source case," *SSP 2011, Nice*, June 2011.
- [39] A.B. Gershman, P. Stoica, M. Pesavento, E.G. Larsson, "Stochastic Cramer-Rao bound for direction estimation in unknown noise fields," *IEE Proc.-Radar Sonar Navig.*, vol. 149, no. 1, pp. 2-8, Feb. 2002.
- [40] A. Gorokhov, Y. Abramovich, and J.F. Böhme, "Unified analysis of DOA estimation algorithms for covariance matrix transforms," *Signal Processing*, vol. 55, pp. 107-115, 1996.
- [41] F. Haddadi, M.M. Mohammadi, M.M. Nayebi, and M.R. Aref, "Statistical performance analysis of MDL source enumeration in array processing," *IEEE Trans. Signal Process.*, vol. 58, no. 1, pp. 452-457, May 2010.
- [42] M. Hawkes and A. Nehorai, "Acoustic vector-sensor beamforming and Capon direction estimation," *IEEE Trans. Signal Process.*, vol. 46, no. 9, pp. 2291-2304, Sep. 1998.
- [43] A.G. Jaffer, "Maximum likelihood direction finding of stochastic sources: a separable solution," in *Proc. ICASSP New York*, pp. 2893-2896, 11-14 April 1988.
- [44] M. Jansson, B. Göransson, and B. Ottersten, "Subspace method for direction of arrival estimation of uncorrelated emitter signals," *IEEE Trans. Signal Process.*, vol. 47, no. 4, pp. 945-956, April 1999.
- [45] B.A. Johnson, Y.I. Abramovich, and X. Mestre, "MUSIC, G-MUSIC, and Maximum-Likelihood Performance Breakdown," *IEEE Trans. Signal Process.*, vol. 56, no. 8, pp. 3944-3958, August 2008.
- [46] T. Kato, *Perturbation Theory for Linear Operators*, Springer-Verlag, Berlin, 1995.
- [47] M. Kaveh and A.J. Barabell, "The statistical performance of the MUSIC and the Minimum-Norm algorithms in resolving plane waves in noise," *IEEE Trans. on ASSP*, vol. 34, no. 2, pp. 331-341, April 1986.
- [48] M. Kaveh, H. Wang, and H. Hung, "On the theoretic performance of a class of estimators of the number of narrow-band sources," *IEEE Trans. on ASSP*, vol. 35, no. 9, pp. 1350-1352, Sept. 1987.
- [49] M. Kaveh and H. Wang, "Threshold properties of narrowband signal subspace array processing methods," in *Advances in Spectrum Analysis and Array Processing*, S. Haykin, Prentice-Hall, pp. 173-220, vol. 2, 1991.
- [50] S.M. Kay, *Fundamentals of Statistical Signal Processing, Detection Theory, vol.II*, Prentice-Hall, 1998.
- [51] H. Krim, P. Forster, and G. Proakis, "Operator approach to performance analysis of root-MUSIC and root-min-norm," *IEEE Trans. Signal Process.*, vol. 40, no. 7, pp. 1687-1696, July 1992.
- [52] M. Lavielle, E. Moulines and J.F. Cardoso, "A maximum likelihood solution to DOA estimation for discrete sources," *Proc. Seventh IEEE Workshop on SP*, pp. 349-353, 1994.
- [53] H.B. Lee and M.S. Wengrovitz, "Resolution threshold of beamspace MUSIC for two closely spaced emitters," *IEEE Trans. ASSP*, vol. 38, no. 9, pp. 1445-1559, Sep. 1990.
- [54] H.B. Lee and M.S. Wengrovitz, "Statistical characterization of the MUSIC algorithm null spectrum," *IEEE Trans. Signal Process.*, vol. 39, no. 6, pp. 1333-1347, June 1991.
- [55] H.B. Lee, "The Cramer-Rao Bound on Frequency Estimates of Signals Closely Spaced in Frequency," *IEEE Trans. Signal Process.*, vol. 40, no. 6, pp. 1508-1517, June 1992
- [56] E.L. Lehmann, *Elements of large-Sample Theory*, Springer-Verlag, New-York, 1999.
- [57] F. Li and R.J. Vaccaro, "Unified analysis for DOA estimation algorithms in array signal processing," *Signal processing*, vol. 25, no. 2, pp. 147-169, Nov. 1991.

- [58] F. Li, H. Liu, and R.J. Vaccaro, "Performance analysis for DOA estimation algorithms: unification, simplification, and observations," *IEEE Trans. on Aerospace and Electronics Systems*, vol. 29, no. 4, pp. 1170-1184, Oct. 1993.
- [59] Z. Liu and A. Nehorai, "Statistical angular resolution limit for point sources," *IEEE Trans. Signal Process.*, vol. 55, no. 11, pp. 5521-5527, Nov. 2007.
- [60] R.R. Nadakuditi and A. Edelman, "Sample eigenvalue based detection of high-dimensional signals in white noise using relatively few samples," *IEEE Trans. Signal Process.*, vol. 56, no. 17, pp. 2625-2638, July 2008.
- [61] A. Nehorai, D. Starer, and P. Stoica, "Direction of arrival estimation with multipath and few snapshots," *Circuits, Syst. Signals, Proc.*, vol. 10, no. 3, pp. 327-342, 1991.
- [62] A. Nehorai and E. Paldi, "Vector-sensor array processing for electromagnetic source localization," *IEEE Trans. Signal Process.*, vol. 42, no. 2, pp. 376-398, Feb. 1994.
- [63] B. Ottersten, M. Viberg, and T. Kailath, "Analysis of subspace fitting and ML techniques for parameter estimation from sensor array data," *IEEE Trans. Signal Process.*, vol. 40, no. 3, pp. 590-599, March 1992.
- [64] B. Ottersten, M. Viberg, P. Stoica, and A. Nehorai, "Exact and large sample maximum likelihood techniques for parameter estimation and detection in array processing," in Haykin, Litva and Shepherd, editors, *Radar Array Processing*, pp. 99-151, Springer-Verlag, Berlin, 1993.
- [65] M. Pesavento and A.B. Gershman, "Maximum-likelihood direction of arrival estimation in the presence of unknown nonuniform noise," *IEEE Trans. Signal Process.*, vol. 49, no. 7, pp. 1310-1324, July 2001.
- [66] S.U. Pillai and G.H. Kwon, "Performance analysis of MUSIC-type high resolution estimators for direction finding in correlated and coherent scenes," *IEEE Trans. on ASSP*, vol. 37, no. 8, pp. 1176-1189, August 1989.
- [67] B. Porat and B. Friedlander, "Analysis of the asymptotic relative efficiency of the MUSIC algorithm," *IEEE Trans. ASSP*, vol. 36, no. 4, pp. 532-544, April 1988.
- [68] B. Porat and B. Friedlander, "Performance analysis of parameter estimation algorithms based on high-order moments," *Int. J. Adaptive Contr. Signal Process.*, vol. 3, pp. 191-229, 1989.
- [69] B. Porat and B. Friedlander, "Direction finding algorithms based on higher order statistics," *IEEE Trans. Signal Process.*, vol. 39, no. 9, pp. 2016-2024, Sept. 1991.
- [70] A. Renaux, P. Forster, E. Chaumette, and P. Larzabal, "On the high SNR conditional maximum likelihood estimator full statistical characterization," *IEEE Trans. Signal Process.*, vol. 54, no. 12, pp. 4840-4843, Dec. 2006.
- [71] A. Renaux, P. Forster, E. Boyer, and P. Larzabal, "Unconditional maximum likelihood performance at finite number of samples and high signal to noise ratio," *IEEE Trans. Signal Process.*, vol. 55, no. 5, pp. 2358-2364, May 2007.
- [72] C.D. Richmond, "Capon algorithm mean-squared error threshold SNR prediction and probability of resolution," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 2748-2764, August 2005.
- [73] J. Rissanen, "Modeling by shortest data description," *Artomatica*, vol. 14, pp. 465-471, 1978.
- [74] P.J. Schreier and L.L. Scharf, *Statistical Signal Processing of Complex-valued Data - the Theory of Improper and Noncircular Signals*, Cambridge University Press, 2010.
- [75] R.J. Serfling, *Approximation theorems of mathematical statistics*, John Wiley and Sons, 1980.
- [76] K.C. Sharman, S.T. Durrani, "Resolving power of signal subspace methods for finite data lengths," in *Proc. ICASSP*, Tampa, Florida, April 1985.
- [77] D. Slepian, "Estimation of signal parameters in the presence of noise," *Trans. IRE Prof. Group Inform. Theory* PG IT-3, pp. 68-89, 1954.
- [78] S.T. Smith, "Statistical resolution limits and the complexified Cramer-Rao bound," *IEEE Trans. Signal Process.*, vol. 53, no. 5, pp. 1597-1609, May 2005.
- [79] J. Sorelius, R.L. Moses, T. Söderström and A.L. Swindlehurst, "Effects of nonzero bandwidth on direction of arrival estimators in array processing," *IEE Proc. Radar, Sonar Navig.*, vol. 145, no. 6, pp. 317-324, Dec. 1998.
- [80] P. Stoica and A. Nehorai, "MUSIC, Maximum likelihood, and Cramer-Rao Bound," *IEEE Trans. ASSP*, vol. 37, no. 5, pp. 720-741, May 1989.
- [81] P. Stoica and A. Nehorai, "Performance study of conditional and unconditional direction of arrival estimation," *IEEE Trans. ASSP*, vol. 38, no. 10, pp. 1783-1795, Oct. 1990.
- [82] P. Stoica and A. Nehorai, "MUSIC, maximum likelihood, and Cramer-Rao bound: Further results and comparisons," *IEEE Trans. ASSP*, vol. 38, no. 12, pp. 2140-2150, Dec. 1990.
- [83] P. Stoica and R.L. Moses, "On biased estimators and the unbiased Cramer-Rao lower bound," *Signal Processing*, vol. 21, pp. 349-350, 1990.
- [84] P. Stoica and N. Nehorai, "On the concentrated stochastic likelihood function in array processing," *Circuit, Syst. Signal Process.*, vol. 14, pp. 669-674, Sept. 1995.
- [85] P. Stoica, B. Ottersten, M. Viberg, and R.L. Moses, "Maximum likelihood array processing for stochastic coherent sources," *IEEE Trans. Signal Process.*, vol. 44, no 1, pp. 96-105, January 1996.
- [86] P. Stoica and R. Moses, *Introduction to spectral analysis*, Prentice hall, Inc. 1997.

- [87] P. Stoica, A.G. Larsson, and A.B. Gershman, "The stochastic CRB for array processing: a textbook derivation," *IEEE Signal Process. letters*, vol. 8, no. 5, pp. 148-150, May 2001.
- [88] T.W.F. Stroud, "Fixed alternatives and Wald's formulation of the noncentral asymptotic behavior of the likelihood ratio statistics," *Annals of Mathematical Statistics*, vol. 43, no. 2, pp. 447-454, 1972.
- [89] A. Stuart and J.K. Ord, *Advanced Theory of Statistics*, fifth edition, vol.2, Edward Arnold, 1991.
- [90] A.L. Swindlehurst and T. Kailath, "A performance analysis of subspace-based methods in the presence of model errors, Part I: The MUSIC algorithm," *IEEE Trans. Signal Process.*, vol. 40, no 7, pp. 1758-1773, July 1992.
- [91] C. Vaidyanathan and K.M. Buckley, "Performance analysis of the MVDR spatial spectrum estimator," *IEEE Trans. Signal Process.*, vol. 43, no. 6, pp. 1427-1437, June 1995.
- [92] P. Vallet, P. Loubaton, and X. Mestre, "Improved subspace estimation for multivariate observations of high dimension: the deterministic signal case," *IEEE Trans. Inform. Theory*, vol. 58, no. 2, pp. 1002-3234, Feb. 2012.
- [93] M. Viberg and B. Ottersten, "Sensor array signal processing based on subspace fitting," *IEEE Trans. ASSP*, vol. 39, no. 5, pp. 1110-1121, May 1991.
- [94] M. Viberg, "Sensitivity of parametric direction finding to colored noise fields and undermodeling," *Signal Processing*, vol. 34, no. 2, pp. 207-222, Nov. 1993.
- [95] M. Viberg and A.L. Swindlehurst, "Analysis of the combined effects of finite samples and model errors on array processing performance," *IEEE Trans. Signal Process.*, vol. 42, no. 11, pp. 3073-3083, Nov. 1994.
- [96] M. Viberg, B. Ottersten, and A. Nehorai, "Performance analysis of direction finding with large arrays and finite data," *IEEE Trans. Signal Process.*, vol. 43, no. 2, pp. 469-477, Feb. 1995.
- [97] D.T. Vu, A. Renaux, R. Boyer, and S. Marcos, "Closed-form expression of the Weiss-Weinstein bound for 3D source localization: The conditional case," In *Proc. SAM*, Jerusalem, Israel, Oct. 2010.
- [98] H. Wang and M. Kaveh, "On the performance of signal subspace processing - Part I: narrow-band systems," *IEEE Trans. ASSP*, vol. 34, no. 5, pp. 1201-1209, Oct. 1986.
- [99] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Trans. ASSP*, vol. 33, no. 2, pp. 387-392, April 1985.
- [100] M. Wax and I. Ziskind, "On unique localization of multiple sources by passive sensor arrays," *IEEE Trans. ASSP*, vol. 37, no. 7, pp. 996-1000, July 1989.
- [101] M. Wax and I. Ziskind, "Detection of the Number of Coherent Signals by the MDL Principle," *IEEE Trans. ASSP*, vol. 37, no. 8, pp. 1190-1196, August 1989.
- [102] G.A. Young and R.L. Smith, *Essentials of Statistical Inference*, Cambridge Series in Statistical and Probabilistic Mathematics, 2005.
- [103] W. Xu and K.M. Buckley, "Bias analysis of the MUSIC location estimator," *IEEE Trans. Signal Process.*, vol. 40, no. 10, pp. 2559-2569, Oct. 1992.
- [104] W. Xu and M Kaveh, "Analysis of the performance and sensitivity of eigendecomposition-based detectors," *IEEE Trans. Signal Process.*, vol. 43, no. 6, pp. 1413-1426, June 1995.
- [105] H. Ye and R.D. Degroat, "Maximum likelihood DOA estimation and asymptotic Cramer-Rao bounds for additive unknown colored noise," *IEEE Trans. Signal Process.*, vol. 43, no. 4, pp. 938-949, April 1995.
- [106] L. C. Zhao, P. R. Krishnaiah, and Z. D. Bai, "On detection of the number of signals in the presence of white noise," *J. Multivariate Anal.*, vol. 20, no. 1, pp. 1-20, Jan. 1986.
- [107] Q.T. Zhang, "Probability of resolution of the MUSIC algorithm," *IEEE Trans. Signal Process.*, vol. 43, no. 4, pp. 978-987, April 1995.
- [108] Q.T. Zhang, "A statistical resolution theory of the beamformer-based spatial spectrum for determining the directions of signals in white noise," *IEEE Trans. on ASSP*, vol. 43, no. 8, pp. 1867-1873, August 1995.
- [109] C. Zhou, F. Haber, and D.L. Jaggard, "A resolution measure for the MUSIC algorithm and its application to plane wave arrivals contaminated by coherent interference," *IEEE Trans. on Signal Process.*, vol. 39, no. 2, pp. 454-463, Feb. 1991.