Postprint of: Meller M. and Stawiarski K., On DoA estimation for rotating arrays using stochastic maximum likelihood approach, in IEEE Transactions on Signal Processing, DOI: 10.1109/TSP.2020.3022207

On DoA estimation for rotating arrays using stochastic maximum likelihood approach

Michał Meller, Kamil Stawiarski

***Abstract*—The flexibility needed to construct DoA estimators that can be used with rotating arrays subject to rapid variations of the signal frequency is offered by the stochastic maximum likelihood approach. Using a combination of analytic methods and Monte Carlo simulations, we show that for low and moderate source correlations the stochastic maximum likelihood estimator that assumes noncorrelated sources has accuracy comparable to the estimator that includes the correlation coefficient as one of the parameters. We propose several fast approximations of the stochastic maximum likelihood estimator and compare their accuracy with the Crámer-Rao lower bound. We also discuss the model order selection problem for the binary- and multiple-hypotheses cases.**

***Index Terms*—direction of arrival estimation, rotating array, stochastic maximum likelihood method**

I. Introduction

In applications such as radar, sonar, or seismology there arises the need to estimate the direction of arrival (DoA) of waveforms accurately. Early DoA estimation methods often involved mechanical movement of the receive array, and subsequent analysis of the resultant amplitude modulation of the received signal. Solutions belonging to this class include averaging of a rotating array azimuth with magnitudes of the array output serving as weights and different variants of the conical scan, among others [1], [2]. However, such methods were found sensitive to fluctuations in the signal strength, and it was soon realized that one may reach better and more robust performance by employing multiple receive beams or, more generally, multiple array outputs [3]. The monopulse method [1], [3], which compares signals received by two beams, is a classical technique belonging to this class.

Later on, numerous so-called superresolution methods, that offer significantly better performance than the monopulse, particularly when there are multiple closely-spaced sources, were proposed. The superresolution, which remains an active area of research to this date, includes wide range of nonparametric and parametric techniques.

In the first class, one may point to, already classical, methods such as the Capon method [4], its extensions [5], several variants of the MUSIC algorithm [6], [7], and the ESPRIT method [8], among others. Recent contributions to this class include, e.g., the minimum norm method [9] or an iterative estimator derived from the weighted least squares method [10]. Nonpararametric methods often include the source independence assumption and require a full rank estimate of the array output covariance matrix. These features make their application difficult when there is only a handful of snapshots (which may be even as low as one) available. To mitigate these difficulties, techniques such as the diagonal loading or the spatial smoothing [11], the latter of which reduces the array resolution, are used widely.

In the second, i.e., parametric class, two approaches, both derived in the maximum likelihood framework, are predominant. The deterministic, or conditional, maximum likelihood estimator treats the source complex amplitudes as unknown deterministic variables, while the stochastic, or unconditional, maximum likelihood estimator treats them as independent realizations of Gaussian random variables [11]. The stochastic model may also be interpreted as a very simple specialization of a wider multiplicative noise model, the latter of which may model spatially spread sources – see, e.g., [12]–[15] for examples of applications and solutions of the DoA estimation under the multiplicative noise problem. Recently, methods employing the sparse recovery concept gained significant interest due to their high accuracy, resolution, and the capability to adjust to different number of sources [16]–[18]. An important feature of almost all parametric methods lies in the fact that they can work with as little as single snapshot without the need to introduce any ad-hoc modifications that may affect their performance adversely.

Remarkably, almost all superresolution methods were developed under the assumption that the source steering vectors are independent of time, which precludes their direct application in systems that employ rotating arrays. In radar, such systems are still widespread, because they offer 360° coverage at costs that are substantially smaller than costs of systems that employ multiple stationary arrays facing different directions. This feature makes rotating-array systems the preferred choice in all but few most demanding applications, such as the ballistic missile defense.

An important exception to this observation are the works of A. Farina, F. Gini, M. Greco, and their coworkers that deal specifically with rotating arrays – see [19]–[22], among others. These authors assumed a coherent radar system and proposed an estimator based on a variant of the conditional maximum likelihood approach that exploits both the rotation-induced amplitude modulation of the array output signals and the Doppler shifts of target echos. They also studied a simplified version of the estimator that avoids high computational cost of a multidimensional search by replacing it with a sequence of smaller two-dimensional searches. The drawback of these estimators lies in the fact that they require a coherent system

Both authors work at the Department of Automatic Control, Gdańsk University of Technology, Faculty of Electronics, Telecommunications and Computer Science, Narutowicza 11/12, 80-233 Gdańsk, Poland and at PIT-RADWAR S.A., Poligonowa 30, 04-051, Warsaw, Poland. e-mails: michal.meller@pitradwar.com, kamil.stawiarski@pitradwar.com

which employs a constant frequency across entire dwell over target, which means that they might be prone to fluctuations of source strength.

Recently, we proposed a nonparametric estimator tailored for rotating arrays that employs the minimum variance distortionless response principle [23]. While this estimator does not require system coherence and is robust to fluctuations, it still has the requirement of constant frequency. Moreover, its computational cost can be quite high.

In this paper, we study DoA estimation for rotating arrays using the stochastic maximum likelihood approach. Unlike the conditional maximum likelihood approach of [19]–[22], the stochastic approach does not require coherence and can easily cope with varying frequency. Additionally, it is more robust to fluctuations of source strengths. On the other hand, in its basic form, the method has rather high computational complexity. The contributions of this paper are the following ones: 1. The introduction of an unconditional model that does not require steering vectors to be independent of time. 2. The sensitivity analysis of the resultant stochastic maximum likelihood estimator. 3. The proposal of three simplified versions of the estimator, which all require substantially less computational power than the basic approach. 4. The verification of analytic results and of the performance of the proposed simplified estimators using Monte Carlo simulations.

The remaining parts of the paper are organized as follows. Section II presents the signal model and the formulation of the problem. Section III studies the stochastic maximum likelihood estimator that arises from the problem formulation. Using analytic methods, we investigate its sensitivity to small errors in estimating the variance parameters and to nonzero source correlation. In Section IV, several simplified, suboptimal, versions of the approach are proposed. Section V presents the results of Monte Carlo simulations. Section VI concludes the paper.

II. Signal model and problem formulation

Let $n = 0, 1, \ldots, N-1$ denote discrete, dimensionless time, and let $\phi_{\mathrm{a},n}$ denote the azimuth of an array at time n. The output of the array at time n, denoted $\boldsymbol{y}_n$, is an M-variate ($M \geq 2$) complex vector that is assumed to originate from the following model

$$\boldsymbol{y}_n = \sum_{k=1}^{K} s_{k,n} \boldsymbol{a}_n(\Delta_{k,n}) + \boldsymbol{v}_n \ , \tag{1}$$

where $K \geq 1$ denotes the number of sources with azimuth angles $\phi_1, \phi_2, \ldots, \phi_K$, $s_{k,n}$ is the scalar complex-valued amplitude of k-th source at time n,

$$\Delta_{k,n} = \phi_k - \phi_{\mathrm{a},n}$$

is the angular deviation of k-th source from the array boresight at time n, $\boldsymbol{a}_n(\Delta\phi_{k,n})$ is the corresponding array steering vector, and $\boldsymbol{v}_n$ is the measurement noise.

We will assume that

(A1) The sources are independent from each other and noise

$$\mathrm{E}[s_{j,m} s^*_{k,n}] = 0 \qquad \forall_{j,k,m,n}$$
$$\mathrm{E}[s_{k,n} \boldsymbol{v}^{\mathrm{H}}_m] = \boldsymbol{0}^{\mathrm{H}} \qquad \forall_{k,m,n}$$

where $s^*_{k,n}$ and $\boldsymbol{v}^{\mathrm{H}}_m$ denote the conjugate of $s_{k,n}$ and the conjugate-transpose of $\boldsymbol{v}_m$, respectively.

(A2) The sequences $\{s_{k,0}, s_{k,1}, \ldots, s_{k,N-1}\}$, $k = 1, 2, \ldots, K$, form zero-mean stationary white complex circular Gaussian random processes with unknown variances $\sigma^2_{s,k}$, $s_{k,n} \sim \mathcal{CN}(0, \sigma^2_{s,k})$, $k = 1, 2, \ldots, K$.

(A3) The measurement noise $\{\boldsymbol{v}_n\}$ is a zero-mean stationary complex circular Gaussian white noise with covariance matrix $\sigma^2_v \boldsymbol{I}$, where $\boldsymbol{I}$ denotes an identity matrix and σ^2_v is unknown, $\boldsymbol{v}_n \sim \mathcal{CN}(\boldsymbol{0}, \sigma^2_v \boldsymbol{I})$,

The presence of multiplicative noise affecting $\boldsymbol{a}_n(\Delta\phi_{k,n})$ is not assumed.

We seek estimators of unknown source angles $\phi_1, \ldots, \phi_K$ that, preferably, have modest computational complexity, which would facilitate their implementation in real-time systems. Moreover, we generally take into account that the size of the observation vector $\boldsymbol{y}_n$ might be small, e.g., $M = 2$, which occurs, among others, if the array design is based on the principles of the monopulse method [1]. The typical monopulse solution, however, employs only one snapshot, which means that the information carried in the envelopes of the sum and difference signals is not processed optimally [21]. Our aim is to deliver an estimator that will offer considerably better performance than the monopulse method, particularly when there are multiple closely spaced sources.

Remark 1: There are several important difference between the rotating and the stationary array cases. First, in the former the antenna rotation makes the observation vectors a nonstationary sequence, which excludes the computation of the sample correlation matrix. Since this quantity is required by almost all superresolution methods, almost all such methods are not directly applicable in the rotating array case. The use of these methods is allowed only under slow array rotation, in which case one could regard observations as locally stationary and employ local analysis in the form of, e.g, the "sliding window" or segmentation approaches. Second, in the rotating array case the movement of the array/elementary source beampattern over sources results in amplitude modulation of observations [19]. The presence or absence of this modulation carries important information about the presence of signal sources and, to reach optimal estimation results, should be taken into account in the estimator design [21], [22]. Third, rotating arrays are a special case of moving arrays, which are more robust to ambiguity problems than stationary ones [24].

Remark 2: Comparing the adopted model with, e.g., the rotating-array radar models considered in [19], [21], one may point to two important differences. First, we adopt the stochastic approach, i.e., we treat the amplitudes $s_{k,n}$ as independent stochastic variables. This feature makes the proposed estimator less affected by source fluctuations, albeit the cost of neglecting the Doppler shift information. More importantly, however, the steering vectors in our model depend not only on the source DoA, but also on the time variable n. Models in the form (1) arise naturally in frequency-agile radars, which may change the operating frequency with each pulse to avoid jamming. Such rapid changes of the pulse frequency alter the array manifold and decorrelate the source (target echo) complex amplitudes $s_{k,n}$ due to the resultant changes in the

wavelength and the scintillation phenomenon [1]. Moreover, in this mode of operation, the system is incoherent, so the Doppler shift information is not available and cannot be analyzed to improve the DoA estimation accuracy.

Remark 3: The source independence assumption (A1) may be regarded as the limiting factor of the proposed approach. Indeed, in many applications the sources are partially correlated. The influence of source correlation will be studied in Sections III and V, where we will show that even though the proposed approach suffers from performance degradation that occurs with correlated sources, it does not exhibit the performance breakdown (known from, e.g., the Capon method) and continues to work even with fully correlated sources. Therefore, the proposed estimators remain useful even in applications where the source correlation is significant – at the very least, they could be used to rapidly provide the initial point for a more sophisticated (and more computationally costly) estimator that does not neglect the source correlation.

III. Stochastic maximum likelihood DoA estimator

A. Exact maximum likelihood solution

We start by discussing the exact maximum likelihood solution. Suppose that the number of sources, K, is known. On the basis of (1) and (A1)-(A3), one may arrive at the following form of the log-likelihood function [25]

$$l(\mathcal{Y},\boldsymbol{\theta}) = C - \left[\sum_{n=0}^{N-1} \log\det \boldsymbol{R}_n(\boldsymbol{\theta}) + \boldsymbol{y}_n^{\mathrm{H}}\boldsymbol{R}_n^{-1}(\boldsymbol{\theta})\boldsymbol{y}_n\right] , \quad (2)$$

where $\mathcal{Y} = \{\boldsymbol{y}_0, \boldsymbol{y}_1, \ldots, \boldsymbol{y}_{N-1}\}$ denotes the available observations,

$$\boldsymbol{\theta} = \left[\begin{array}{cccccc} \phi_1 & \ldots & \phi_K & \sigma_{s,1}^2 & \ldots & \sigma_{s,K}^2 \quad \sigma_v^2 \end{array}\right]^{\mathrm{T}}$$

is the vector of the model parameters, C is a constant whose exact form is not important to our considerations, and

$$\boldsymbol{R}_n(\boldsymbol{\theta}) = \sigma_v^2\boldsymbol{I} + \sum_{k=1}^{K}\sigma_{s,k}^2\boldsymbol{a}_n(\Delta_{k,n})\boldsymbol{a}_n^{\mathrm{H}}(\Delta_{k,n}) \quad (3)$$

is the covariance matrix of the observation vector at time n. The maximum likelihood estimator follows immediately

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \; l(\mathcal{Y},\boldsymbol{\theta}) \; . \quad (4)$$

Let us decompose $\boldsymbol{\theta}$ into the DoA subvector and the variance subvector, $\boldsymbol{\theta} = [\boldsymbol{\theta}_\phi \; \boldsymbol{\theta}_\sigma]^{\mathrm{T}}$, where

$$\boldsymbol{\theta}_\phi = \left[\begin{array}{ccc} \phi_1 & \ldots & \phi_K \end{array}\right]^{\mathrm{T}}$$
$$\boldsymbol{\theta}_\sigma = \left[\begin{array}{cccc} \sigma_{s,1}^2 & \ldots & \sigma_{s,K}^2 \quad \sigma_v^2 \end{array}\right]^{\mathrm{T}} .$$

Since our primary interest lies in estimating $\boldsymbol{\theta}_\phi$, we may treat $\boldsymbol{\theta}_\sigma$ as nuisance parameters. Although the negative log-likelihood is generally nonconvex in $\boldsymbol{\theta}_\phi$, it is straightforward to verify that, irrespective of the DoAs in $\boldsymbol{\theta}_\phi$, the negative likelihood function is convex in the variance variables, $\boldsymbol{\theta}_\sigma$, because it is a composition of convex mappings in these variables [26]. This property suggest implementing the maximization of $l(\mathcal{Y},\boldsymbol{\theta})$ as the alternating sequence of "outer", exhaustive, search in $\boldsymbol{\theta}_\phi$, and the "inner" maximization in $\boldsymbol{\theta}_\sigma$ that can take advantage of the convexity,

$$\hat{\boldsymbol{\theta}}_\phi = \arg\max_{\boldsymbol{\theta}_\phi} \; \bar{l}(\mathcal{Y},\boldsymbol{\theta}_\phi) \; , \quad (5)$$

where

$$\bar{l}(\mathcal{Y},\boldsymbol{\theta}_\phi) = l(\mathcal{Y},\boldsymbol{\theta}_\phi,\hat{\boldsymbol{\theta}}_\sigma(\boldsymbol{\theta}_\phi))$$
$$\hat{\boldsymbol{\theta}}_\sigma(\boldsymbol{\theta}_\phi) = \arg\max_{\boldsymbol{\theta}_\sigma} \; l(\mathcal{Y},\boldsymbol{\theta}_\phi,\boldsymbol{\theta}_\sigma)$$

denotes the compressed log-likelihood function.

The problem of the outer search will not be discussed here extensively. We only point that it can be implemented as two-stage search over a relatively coarse K-dimensional grid of DoAs to localize the neighborhood of the global maximum, followed by finer search to localize the optimal DoAs exactly. In the second stage, we recommend employing a derivative-free algorithm to avoid evaluating the derivatives of $\bar{l}(\mathcal{Y},\boldsymbol{\theta}_\phi)$. Such an approach makes a lot of sense if the array manifold is stored in the form of lookup tables, because in this situation the derivatives of steering vectors might be unavailable.

Unlike in the case of stationary array, where a closed-form solution of the inner optimization exists [27], in the discussed case no explicit formula for the optimal values of $\boldsymbol{\theta}_\sigma$ is known to exist. However, due to the convexity of the problem, the convergence of the steepest descent method

$$\hat{\boldsymbol{\theta}}_\sigma \leftarrow \hat{\boldsymbol{\theta}}_\sigma + \alpha\frac{\partial l(\mathcal{Y},\boldsymbol{\theta}_\phi,\boldsymbol{\theta}_\sigma)}{\partial\boldsymbol{\theta}_\sigma} \; , \quad (6)$$

where $\alpha > 0$ is a small gain, is guaranteed. After straightforward manipulations, one may show that the partial derivatives needed to implement the method take the form

$$\frac{\partial l}{\partial\sigma_{s,k}^2} = \sum_{n=0}^{N-1}\left|\boldsymbol{y}_n^{\mathrm{H}}\boldsymbol{R}_n^{-1}(\boldsymbol{\theta}_\phi,\boldsymbol{\theta}_\sigma)\boldsymbol{a}_n(\Delta_{k,n})\right|^2 - \sum_{n=0}^{N-1}\boldsymbol{a}_n^{\mathrm{H}}(\Delta_{k,n})\boldsymbol{R}_n^{-1}(\boldsymbol{\theta}_\phi,\boldsymbol{\theta}_\sigma)\boldsymbol{a}_n(\Delta_{k,n}) \quad (7)$$

$$\frac{\partial l}{\partial\sigma_v^2} = \sum_{n=0}^{N-1}\boldsymbol{y}_n^{\mathrm{H}}\boldsymbol{R}_n^{-2}(\boldsymbol{\theta}_\phi,\boldsymbol{\theta}_\sigma)\boldsymbol{y}_n - \sum_{n=0}^{N-1}\mathrm{tr}\;\boldsymbol{R}_n^{-1}(\boldsymbol{\theta}_\phi,\boldsymbol{\theta}_\sigma) \quad (8)$$

Unfortunately, this approach is computationally expensive and therefore impractical. In the next section, we will present several options, which all involve approximations, that have lower computational complexity. Prior to this, however, we will show that, under mild conditions, the DoA estimates obtained using the maximum likelihood estimator are fairly insensitive to small errors in estimating the optimal values of the variance parameters, which means that the adverse influence of approximations should be tolerable, if not negligible.

B. Sensitivity analysis of the maximum likelihood estimator

Denote by $\boldsymbol{\theta}_0 = [\boldsymbol{\theta}_{\phi,0}^{\mathrm{T}} \; \boldsymbol{\theta}_{\sigma,0}^{\mathrm{T}}]^{\mathrm{T}}$ the true value of the parameter vector $\boldsymbol{\theta}$. One can analyze the behavior of maximum likelihood estimators using the Taylor expansion of the log-likelihood

derivative about $\boldsymbol{\theta}_0$ [28, chap. 2.2]. Expanding the first derivative of $l(\mathcal{Y},\boldsymbol{\theta})$

$$\frac{\partial l(\mathcal{Y},\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \simeq \frac{\partial l(\mathcal{Y},\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} + \frac{\partial^2 l(\mathcal{Y},\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}^{\mathrm{T}}}(\boldsymbol{\theta}-\boldsymbol{\theta}_0)$$

one obtains that the maximum likelihood estimate $\hat{\boldsymbol{\theta}}$, which by definition satisfies

$$\frac{\partial l(\mathcal{Y},\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}} = \mathbf{0}\ ,$$

reads

$$\hat{\boldsymbol{\theta}} \simeq \boldsymbol{\theta}_0 - \left[\frac{\partial^2 l(\mathcal{Y},\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}^{\mathrm{T}}}\right]^{-1} \frac{\partial l(\mathcal{Y},\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \simeq \boldsymbol{\theta}_0 + \boldsymbol{J}^{-1}\frac{\partial l(\mathcal{Y},\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \tag{9}$$

where

$$\boldsymbol{J} = -\mathrm{E}\left[\frac{\partial^2 l(\mathcal{Y},\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}^{\mathrm{T}}}\right] = \begin{bmatrix} \boldsymbol{J}_{\phi\phi} & \boldsymbol{J}_{\phi\sigma} \\ \boldsymbol{J}_{\sigma\phi} & \boldsymbol{J}_{\sigma\sigma} \end{bmatrix} \tag{10}$$

is the Fisher information matrix. It is well know that, for $n \to \infty$, the random variable $\frac{1}{n}\frac{\partial l(\mathcal{Y},\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}$ converges in distribution to a zero-mean Gaussian random variable with covariance matrix $\boldsymbol{J}$ [28, chap. 2.2].

Suppose that a perturbed log-likelihood maximizer in $\boldsymbol{\theta}_\sigma$, $\hat{\boldsymbol{\theta}}_\sigma + \Delta\hat{\boldsymbol{\theta}}_\sigma$ is found. We are interested in evaluating how such an error affects the maximizer in $\boldsymbol{\theta}_\phi$. One can quantify this influence by considering the subvector of $\partial l(\mathcal{Y},\boldsymbol{\theta})/\partial\boldsymbol{\theta}$ that corresponds to $\boldsymbol{\theta}_\phi$, and equating it to zero

$$\begin{aligned}\frac{\partial l(\mathcal{Y},\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}_\phi} &+ \frac{\partial^2 l(\mathcal{Y},\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}_\phi \partial \boldsymbol{\theta}_\phi^{\mathrm{T}}}(\hat{\boldsymbol{\theta}}_\phi + \Delta\hat{\boldsymbol{\theta}}_\phi - \boldsymbol{\theta}_{\phi,0}) \\ &+ \frac{\partial^2 l(\mathcal{Y},\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}_\phi \partial \boldsymbol{\theta}_\sigma^{\mathrm{T}}}(\hat{\boldsymbol{\theta}}_\sigma + \Delta\hat{\boldsymbol{\theta}}_\sigma - \boldsymbol{\theta}_{\sigma,0}) = \mathbf{0}\ .\end{aligned}$$

As a special case, the above equality must hold for $\Delta\hat{\boldsymbol{\theta}}_\phi = \mathbf{0}$, $\Delta\hat{\boldsymbol{\theta}}_\sigma = \mathbf{0}$, which leads us to

$$\frac{\partial^2 l(\mathcal{Y},\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}_\phi \partial \boldsymbol{\theta}_\phi^{\mathrm{T}}}\Delta\hat{\boldsymbol{\theta}}_\phi + \frac{\partial^2 l(\mathcal{Y},\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}_\phi \partial \boldsymbol{\theta}_\sigma^{\mathrm{T}}}\Delta\hat{\boldsymbol{\theta}}_\sigma = \mathbf{0}$$

and therefore we obtain that

$$\begin{aligned}\Delta\hat{\boldsymbol{\theta}}_\phi &= -\left[\frac{\partial^2 l(\mathcal{Y},\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}_\phi \partial \boldsymbol{\theta}_\phi^{\mathrm{T}}}\right]^{-1} \frac{\partial^2 l(\mathcal{Y},\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}_\phi \partial \boldsymbol{\theta}_\sigma^{\mathrm{T}}}\Delta\hat{\boldsymbol{\theta}}_\sigma \\ &\simeq -\boldsymbol{J}_{\phi\phi}^{-1}\boldsymbol{J}_{\phi\sigma}\Delta\hat{\boldsymbol{\theta}}_\sigma\ .\end{aligned} \tag{11}$$

Given that $\boldsymbol{J}_{\phi\phi}$ must be a full rank matrix for $\boldsymbol{J}_{\phi\phi}^{-1}$ to exist, the existence of the influence of $\Delta\hat{\boldsymbol{\theta}}_\sigma$ is decided by $\boldsymbol{J}_{\phi\sigma}$ – if $\boldsymbol{J}_{\phi\sigma}$ is a zero matrix, then (up to the first order terms), the error $\Delta\hat{\boldsymbol{\theta}}_\sigma$ does not affect the estimates of the source DoAs [also note close ties of (11) with the formula for the Crámer-Rao lower bound on DoAs obtained by treating $\boldsymbol{\theta}_\sigma$ as nuisance].

It is straightforward to verify that $\boldsymbol{J}_{\phi\sigma}$ is made of the elements of the form [c.f. (10)]

$$\begin{aligned}&-\mathrm{E}\left[\frac{\partial^2 l(\mathcal{Y},\boldsymbol{\theta}_0)}{\partial \phi_k \partial \sigma_x^2}\right] = \\ &\sum_{n=0}^{N-1} \mathrm{tr}\left(\boldsymbol{R}_n^{-1}(\boldsymbol{\theta}_0)\frac{\partial \boldsymbol{R}_n(\boldsymbol{\theta}_0)}{\partial \phi_k}\boldsymbol{R}_n^{-1}(\boldsymbol{\theta}_0)\frac{\partial \boldsymbol{R}_n(\boldsymbol{\theta}_0)}{\partial \sigma_x^2}\right)\end{aligned}$$

where $k = 1, 2, \ldots, K$, the symbol σ_x^2 corresponds to any of the variances in the model, and [c.f. (3)]

$$\begin{aligned}\frac{\partial \boldsymbol{R}_n(\boldsymbol{\theta}_0)}{\partial \phi_k} &= \sigma_k^2\left[\frac{\partial \boldsymbol{a}_n(\Delta_{k,n})}{\partial \phi_k}\boldsymbol{a}_n^{\mathrm{H}}(\Delta_{k,n}) \right.\\ &\qquad \left. +\ \boldsymbol{a}_n(\Delta_{k,n})\frac{\partial \boldsymbol{a}_n^{\mathrm{H}}(\Delta_{k,n})}{\partial \phi_k}\right] \\ \frac{\partial \boldsymbol{R}_n(\boldsymbol{\theta}_0)}{\partial \sigma_j^2} &= \boldsymbol{a}_n(\Delta_{j,n})\boldsymbol{a}_n^{\mathrm{H}}(\Delta_{j,n}) \\ \frac{\partial \boldsymbol{R}_n(\boldsymbol{\theta}_0)}{\partial \sigma_v^2} &= \boldsymbol{I}\ .\end{aligned}$$

General discussion of properties of $\boldsymbol{J}_{\phi\sigma}$ is difficult. However, for $K = 1$ and N sufficiently large for the above analysis to be valid, one can verify that the sufficient (but not necessary) condition for $\boldsymbol{J}_{\phi\sigma} = \mathbf{0}$ is that the array manifold $\boldsymbol{a}_n(\Delta\phi)$ is a symmetric function and that the sequence of array azimuth angles $\phi_{\mathrm{a},n}$, $n = 0, 1, \ldots, N-1$, also exhibits the symmetry about ϕ_1. Moreover, if the symmetry of array angles $\phi_{\mathrm{a},n}$ about ϕ_1 is not exact, but the array azimuth angles sample the mainlobe of the array beampattern sufficiently densely, then $\boldsymbol{J}_{\phi\sigma} \simeq \mathbf{0}$ and the influence of $\Delta\hat{\boldsymbol{\theta}}_\sigma$ is small. One may expect that similar properties hold for multiple sources, provided that they are sufficiently well separated from each other, which creates a strong incentive to replace the steepest descent approach (6) with faster methods that employ various approximations.

C. Influence of source correlation

Although we already pointed out why in certain applications it is not unrealistic to assume that the sources are not correlated, it is worth to analyze how the maximum likelihood approach (4) reacts to small, but nonzero, source correlation. To this end one may use the framework presented in chapter 5 of [28]. Related material can also be found in [29]. For simplicity, we will discuss the case of two sources, $K = 2$, only. Consider the following model

$$\boldsymbol{y}_n = \boldsymbol{A}_n(\Delta_{1,n},\Delta_{2,n})\boldsymbol{s}_n + \boldsymbol{v}_n\ , \tag{12}$$

where

$$\boldsymbol{A}_n(\Delta_{1,n},\Delta_{2,n}) = \begin{bmatrix} \boldsymbol{a}_n(\Delta_{1,n}) & \boldsymbol{a}_n(\Delta_{2,n}) \end{bmatrix}$$

and $\{\boldsymbol{s}_n\}$ is a sequence of i.i.d. bivariate complex circular Gaussian random variables with covariance matrix

$$\mathrm{E}[\boldsymbol{s}_n\boldsymbol{s}_n^{\mathrm{H}}] = \begin{bmatrix} \sigma_{s,1}^2 & \rho\sigma_{s,1}\sigma_{s,2} \\ \rho^*\sigma_{s,1}\sigma_{s,2} & \sigma_{s,2}^2 \end{bmatrix}\ ,$$

where $\rho = \rho_{\mathrm{r}} + j\rho_{\mathrm{i}}$, $|\rho| \le 1$, $|\rho| \approx 0$ is the normalized correlation coefficient of the two sources. We will refer to the model (12) as the wide model. Our basic model (1), which one can obtain from (12) by setting $\rho = 0$, will be called the narrow model.

Denote by $\boldsymbol{\theta}_w = [\theta^{\mathrm{T}}\ \rho_{\mathrm{r}}\ \rho_{\mathrm{i}}]^{\mathrm{T}}$ the vector of parameters of the wide model. Let $\boldsymbol{J}_w$ denote the Fisher information matrix of the wide model, evaluated for $\rho = 0$, and partition it as

$$\boldsymbol{J}_{\mathrm{wide}} = \begin{bmatrix} \boldsymbol{J}_{00} & \boldsymbol{J}_{01} \\ \boldsymbol{J}_{10} & \boldsymbol{J}_{11} \end{bmatrix}$$

where $\boldsymbol{J}_{00} = \boldsymbol{J}$ is the Fisher information matrix of the narrow model (10) and the remaining submatrices result from the inclusion of ρ_r, ρ_i into $\boldsymbol{\theta}_w$. Then, according to [28, chap. 5.4], for small ρ and sufficiently large N the bias and variance of the estimate of a parameter μ obtained using the maximum likelihood estimator based on the narrow model are approximately equal to

$$\begin{aligned}\mathrm{E}[\hat{\mu}_\mathrm{narrow} - \mu_0] &\simeq \boldsymbol{\omega}^\mathrm{T}\boldsymbol{\gamma} \\ \mathrm{E}[\hat{\mu}_\mathrm{narrow} - \mathrm{E}[\hat{\mu}_\mathrm{narrow}]]^2 &\simeq \tau_0^2 \,, \end{aligned} \tag{13}$$

where $\boldsymbol{\gamma} = [\rho_\mathrm{r}\ \rho_\mathrm{i}]^\mathrm{T}$,

$$\begin{aligned}\boldsymbol{\omega} &= \boldsymbol{J}_{10}\boldsymbol{J}_{00}^{-1}\frac{\partial\mu}{\partial\boldsymbol{\theta}} - \frac{\partial\mu}{\partial\boldsymbol{\gamma}} \\ \tau_0^2 &= \left(\frac{\partial\mu}{\partial\boldsymbol{\theta}}\right)^\mathrm{T}\boldsymbol{J}_{00}^{-1}\frac{\partial\mu}{\partial\boldsymbol{\theta}} \,,\end{aligned}$$

and the derivatives are evaluated at $\rho_\mathrm{r} = \rho_\mathrm{i} = 0$. For $\mu = \phi_1$ or ϕ_2, one should employ $\frac{\partial\mu}{\partial\boldsymbol{\theta}} = [1\ 0\ 0\ 0\ 0]^\mathrm{T}$ and $\frac{\partial\mu}{\partial\boldsymbol{\theta}} = [0\ 1\ 0\ 0\ 0]^\mathrm{T}$, respectively. In both cases, $\frac{\partial\mu}{\partial\boldsymbol{\gamma}} = [0\ 0]^\mathrm{T}$.

The wide estimator is approximately unbiased with variance

$$\mathrm{E}[\hat{\mu}_\mathrm{wide} - \mathrm{E}[\hat{\mu}_\mathrm{wide}]]^2 = \tau_0^2 + \boldsymbol{\omega}^\mathrm{T}\boldsymbol{K}\boldsymbol{\omega} \,, \tag{14}$$

where $\boldsymbol{K} = (\boldsymbol{J}_{11} - \boldsymbol{J}_{10}\boldsymbol{J}_{00}^{-1}\boldsymbol{J}_{01})^{-1}$ [28, chap 5.4]. It is straightforward to verify by inspection that (14) is simply the Crámer-Rao lower bound for the wide model at $\rho_\mathrm{r} = \rho_\mathrm{i} = 0$.

For both estimators, one can obtain their mean-squared error by adding squared bias and variance

$$\mathrm{MSE}_\mathrm{narrow} \simeq \tau_0^2 + \boldsymbol{\omega}^\mathrm{T}\boldsymbol{\gamma}\boldsymbol{\gamma}^\mathrm{T}\boldsymbol{\omega} \tag{15}$$

$$\mathrm{MSE}_\mathrm{wide} \simeq \tau_0^2 + \boldsymbol{\omega}^\mathrm{T}\boldsymbol{K}\boldsymbol{\omega} \,. \tag{16}$$

Observe that the narrow estimator is always better than the wide one in terms of variance [c.f. (13), (14)], and that it may be better in terms of the MSE, provided that its squared bias is smaller than $\boldsymbol{\omega}^\mathrm{T}\boldsymbol{K}\boldsymbol{\omega}$. Since the bias depends linearly on γ, such performance advantage is, in fact, guaranteed for ρ that is sufficiently close to zero or for such ρ that results in γ being orthogonal to $\boldsymbol{\omega}$. The exact level of source correlation where the accuracy of the two approaches becomes equal depends on many factors, such as the array manifold, source separation, and the number of observations, among others. While it is difficult to draw general conclusions, one may use the above described approach to quickly quantify the tradeoffs associated with the choice of the estimator for a particular combination of system parameters and requirements. In Section V, we will present an example of such analysis in a simple system that employs the monopulse array configuration.

IV. Fast approximate maximum likelihood estimators

In this section, we present fast approximate solutions to the problem of the inner optimization over $\boldsymbol{\theta}_\sigma$. We start with the case $K = 1$, where we present a revised version of the estimator from [30] and explain basic underlying ideas. Then we discuss the generic case of arbitrary K. We also propose a test that allows one to choose the value of K if the number of sources is unknown.

A. Approximate solution for $K = 1$

It is well known that in the stationary case the optimal estimates of source and noise variances can be found using projections on signal and noise subspaces, respectively [27]. We will adopt the same approach, even though with our model it does not lead to optimal estimates. Suppose that $K = 1$ and denote by

$$\begin{aligned}\boldsymbol{Q}_n(\boldsymbol{\theta}_\phi) &= \boldsymbol{I} - \frac{\boldsymbol{a}_n(\Delta_{1,n})\boldsymbol{a}_n^\mathrm{H}(\Delta_{1,n})}{\boldsymbol{a}_n^\mathrm{H}(\Delta_{1,n})\boldsymbol{a}_n(\Delta_{1,n})} \\ \boldsymbol{P}_n(\boldsymbol{\theta}_\phi) &= \frac{\boldsymbol{a}_n(\Delta_{1,n})\boldsymbol{a}_n^\mathrm{H}(\Delta_{1,n})}{\boldsymbol{a}_n^\mathrm{H}(\Delta_{1,n})\boldsymbol{a}_n(\Delta_{1,n})}\end{aligned}$$

the matrices that project the n-th observation on the noise and signal subspaces corresponding to the assumed source DoAs $\boldsymbol{\theta}_\phi$. It is straightforward to show that if the assumed DoAs match their true values, $\boldsymbol{\theta}_\phi = \boldsymbol{\theta}_{\phi,0}$, it holds that [c.f. (1)]

$$\mathrm{E}\left[\sum_{n=0}^{N-1}\boldsymbol{y}_n^\mathrm{H}\boldsymbol{Q}_n(\boldsymbol{\theta}_{\phi,0})\boldsymbol{y}_n\right] = N(M-1)\sigma_v^2 \,, \tag{17}$$

Since it is not unreasonable to assume that $N \gg 1$, one is allowed to employ the following approximation

$$\frac{1}{N}\mathrm{E}\left[\sum_{n=0}^{N-1}\boldsymbol{y}_n^\mathrm{H}\boldsymbol{Q}_n(\boldsymbol{\theta}_{\phi,0})\boldsymbol{y}_n\right] \simeq \frac{1}{N}\sum_{n=0}^{N-1}\boldsymbol{y}_n^\mathrm{H}\boldsymbol{Q}_n(\boldsymbol{\theta}_{\phi,0})\boldsymbol{y}_n \tag{18}$$

which leads to the following estimator of the noise variance [30]

$$\hat{\sigma}_v^2(\boldsymbol{\theta}_\phi) = \frac{\sum_{n=0}^{N-1}\boldsymbol{y}_n^\mathrm{H}\boldsymbol{Q}_n(\boldsymbol{\theta}_\phi)\boldsymbol{y}_n}{(M-1)N} \,. \tag{19}$$

Similarly, it holds that [again, c.f. (1)]

$$\mathrm{E}\left[\sum_{n=0}^{N-1}\boldsymbol{y}_n^\mathrm{H}\boldsymbol{P}_n(\boldsymbol{\theta}_{\phi,0})\boldsymbol{y}_n\right] = N\sigma_v^2 + \sigma_{s,1}^2\sum_{n=0}^{N-1}\|\boldsymbol{a}_n(\Delta_{1,n})\|^2 \tag{20}$$

Based on the reasoning analogous to the one used for σ_v^2, one arrives at following estimator of the source variance

$$\hat{\sigma}_{s,1}^2(\boldsymbol{\theta}_\phi) = \frac{\sum_{n=0}^{N-1}\boldsymbol{y}_n^\mathrm{H}\boldsymbol{P}_n(\boldsymbol{\theta}_\phi)\boldsymbol{y}_n - N\hat{\sigma}_v^2(\boldsymbol{\theta}_\phi)}{\sum_{n=0}^{N-1}\|\boldsymbol{a}_n(\Delta_{1,n})\|^2} \,. \tag{21}$$

Remark: If the array is assumed stopped and the steering vectors independent of n, the above method reduces to the optimal separable solution presented in [27]. In the general case, however, the variance estimates obtained using (19) and (21) are not optimal, although they are consistent.

B. Approximate solutions for arbitrary known K

The proposed solution for an arbitrary number of targets also involves projections. It is straightforward to show that when the assumed DoAs match their true values

$$\begin{aligned}\mathrm{E}\left[\sum_{n=0}^{N-1}w(\Delta_{k,n})\boldsymbol{y}_n^\mathrm{H}\boldsymbol{P}_{k,n}(\boldsymbol{\theta}_{\phi,0})\boldsymbol{y}_n\right] &= \sigma_v^2\sum_{n=0}^{N-1}w(\Delta_{k,n}) \\ &+ \sum_{j=1}^{K}\sigma_{s,j}^2\sum_{n=0}^{N-1}w(\Delta_{k,n})\boldsymbol{a}_n^\mathrm{H}(\Delta_{j,n})\boldsymbol{P}_{k,n}(\boldsymbol{\theta}_{\phi,0})\boldsymbol{a}_n(\Delta_{j,n}) \,,\end{aligned} \tag{22}$$

where $k = 1, 2, \ldots, K$,

$$\boldsymbol{P}_{k,n}(\boldsymbol{\theta}_\phi) = \frac{\boldsymbol{a}_n(\Delta_{k,n})\boldsymbol{a}_n^{\mathrm{H}}(\Delta_{k,n})}{\boldsymbol{a}_n^{\mathrm{H}}(\Delta_{k,n})\boldsymbol{a}_n(\Delta_{k,n})}$$

and $w(\Delta_{k,n})$ is a weighting function, decaying towards zero as $|\Delta_{k,n}|$ increases, whose purpose will be to improve the stability of the estimator.

As in the case of $K = 1$, for $N \gg 1$ one is allowed to drop the expectation on the left side of the above equation. Gathering the resultant equations for $k = 1, 2, \ldots, K$, one may arrive at the following system of equations

$$\boldsymbol{S}(\boldsymbol{\theta}_{\phi,0})\boldsymbol{\sigma}_s \simeq \boldsymbol{u}(\boldsymbol{\theta}_{\phi,0}, \sigma_v^2) \;, \tag{23}$$

where $\boldsymbol{\sigma}_s = [\sigma_{s,1}^2 \; \sigma_{s,2}^2 \; \ldots \; \sigma_{s,K}^2]^{\mathrm{T}}$ and

$$\begin{aligned}
[\boldsymbol{S}(\boldsymbol{\theta}_\phi)]_{ij} &= \sum_{n=0}^{N-1} w(\Delta_{i,n})\boldsymbol{a}_n^{\mathrm{H}}(\Delta_{j,n})\boldsymbol{P}_{i,n}(\boldsymbol{\theta}_\phi)\boldsymbol{a}_n(\Delta_{j,n}) \\
&= \begin{cases} \sum_{n=0}^{N-1} w(\Delta_{i,n})\boldsymbol{a}_n^{\mathrm{H}}(\Delta_{i,n})\boldsymbol{a}_n(\Delta_{i,n}) & \text{for} \quad i = j \\ \sum_{n=0}^{N-1} w(\Delta_{i,n}) \frac{|\boldsymbol{a}_n^{\mathrm{H}}(\Delta_{i,n})\boldsymbol{a}_n(\Delta_{j,n})|^2}{\boldsymbol{a}_n^{\mathrm{H}}(\Delta_{i,n})\boldsymbol{a}_n(\Delta_{i,n})} & \text{for} \quad i \neq j \end{cases} \\
[\boldsymbol{u}(\boldsymbol{\theta}_\phi, \sigma_v^2)]_i &= \sum_{n=0}^{N-1} w(\Delta_{i,n}) \left[\boldsymbol{y}_n^{\mathrm{H}}\boldsymbol{P}_{i,n}(\boldsymbol{\theta}_\phi)\boldsymbol{y}_n - \sigma_v^2\right] \;. \qquad (24)
\end{aligned}$$

Suppose that a fairly accurate estimate of the noise variance, $\hat{\sigma}_v^2(\boldsymbol{\theta}_\phi)$, is available. Replacing σ_v^2 in $\boldsymbol{u}(\boldsymbol{\theta}_\phi)$ with $\hat{\sigma}_v^2(\boldsymbol{\theta}_\phi)$, one arrives at the following estimator of the source variances

$$\hat{\boldsymbol{\sigma}}_s(\boldsymbol{\theta}_\phi, \hat{\sigma_v}^2) = \boldsymbol{S}^{-1}(\boldsymbol{\theta}_\phi)\boldsymbol{u}(\boldsymbol{\theta}_\phi, \hat{\sigma_v}^2(\boldsymbol{\theta}_\phi)) \;. \tag{25}$$

We are now left with the problem of evaluating $\hat{\sigma}_v^2(\boldsymbol{\theta}_\phi)$. If $M > K$, one may employ the projection on the noise subspace

$$\hat{\sigma}_v^2(\boldsymbol{\theta}_\phi) = \frac{\sum_{n=0}^{N-1} \boldsymbol{y}_n^{\mathrm{H}}\boldsymbol{Q}_n(\boldsymbol{\theta}_\phi)\boldsymbol{y}_n}{(M-K)N} \;, \tag{26}$$

where $\boldsymbol{Q}_n(\boldsymbol{\theta}_\phi)$ is the noise subspace projection matrix, i.e., a matrix that projects $\boldsymbol{y}_n$ on a subspace orthogonal to $\boldsymbol{a}_n(\Delta_{k,n})$, $k = 1, 2, \ldots, K$. We will refer to this solution as ASML-1 (approximate stochastic maximum likelihood 1). Note that, for $K = 1$, this algorithm reduces to (19)-(21).

We are more interested, however, in the more challenging case of $K \geq M$, when (26) fails because of the division by zero. To cope with this difficulty, we propose two more solutions, referred to as ASML-2 and ASML-3, respectively.

ASML-2 algorithm is based on the straightforward observation that

$$\mathrm{E}\left[\sum_{n=0}^{N-1} \|\boldsymbol{y}_n\|^2\right] = MN\sigma_v^2 + \sum_{k=1}^{K} \sigma_{s,k}^2 \sum_{n=0}^{N-1} \|\boldsymbol{a}_n(\Delta_{k,n})\|^2 \;, \tag{27}$$

which leads to the estimator of σ_v^2 based on the excess energy

$$\hat{\sigma}_v^2(\boldsymbol{\theta}_\phi) = \frac{1}{NM}\left[\sum_{n=0}^{N-1} \|\boldsymbol{y}_n\|^2 - \sum_{k=1}^{K} \hat{\sigma}_{s,k}^2(\boldsymbol{\theta}_\phi) \sum_{n=0}^{N-1} \|\boldsymbol{a}_n(\Delta_{k,n})\|^2\right] \;. \tag{28}$$

Note that, since (28) depends on $\hat{\sigma}_{s,1}^2(\boldsymbol{\theta}_\phi), \ldots, \hat{\sigma}_{s,K}^2(\boldsymbol{\theta}_\phi)$ and (25) depends on $\hat{\sigma}_v^2(\boldsymbol{\theta}_\phi)$, the implementation of the estimator requires one to alternate the descent step (28) with the reestimation of the source variances $\sigma_{s,1}^2, \ldots, \sigma_{s,K}^2$ using (25). To detect the convergence of this scheme, one may test if the one step changes of all variances fall below prespecified thresholds.

ASML-3 employs the steepest descent approach, where we plug the estimates $\hat{\sigma}_{s,1}^2(\boldsymbol{\theta}_\phi), \ldots, \hat{\sigma}_{s,K}^2(\boldsymbol{\theta}_\phi)$ into the iteration

$$\hat{\sigma}_v^2(\boldsymbol{\theta}_\phi) \leftarrow \hat{\sigma}_v^2(\boldsymbol{\theta}_\phi) + \alpha \frac{\partial l(\mathcal{Y}, \boldsymbol{\theta}_\phi, \hat{\boldsymbol{\theta}}_\sigma(\boldsymbol{\theta}_\phi))}{\partial \sigma_v^2} \;. \tag{29}$$

The initial estimate of $\hat{\sigma}_v^2(\boldsymbol{\theta}_\phi)$ can be based on a priori knowledge of the receiver noise level or using single iteration of ASML-2. Similar to ASML-2, one should alternate (25) and (29) until both recursions converge. Moreover, we propose to use $\alpha = \hat{\sigma}_v^4(\boldsymbol{\theta}_\phi)/NM$, which typically allows the algorithm to converge in about five to ten steps.

C. Summary and discussion

We presented three approximate stochastic maximum likelihood estimators, called ASML-1, ASML-2, and ASML-3.

ASML-1 is the most lightweight option, because it does not involve iterative evaluation of $\hat{\boldsymbol{\theta}}_\sigma(\boldsymbol{\theta}_\phi)$. However, it can be used only if $M > K$, which is somewhat restrictive. We will therefore focus on the ASML-2 and ASML-3 algorithms, which are summarized in Table I, and are both free of such a restriction. Both estimators include basic safeguards that ensure that no element of $\hat{\boldsymbol{\theta}}_\sigma$ is negative and that the noise variance does not fall below a certain minimal level $\sigma_{v,\min}^2$.

ASML-2 and ASML-3 involve the iterative refinement of $\hat{\boldsymbol{\theta}}_\sigma(\boldsymbol{\theta}_\phi)$. However, note that the computational complexity of ASML-2 is quite low, because one may precompute almost all quantities that are needed in steps 2-4, such as entire matrix $\boldsymbol{S}(\boldsymbol{\theta}_\phi)$, the terms $\sum_{n=0}^{N-1} \boldsymbol{y}_n^{\mathrm{H}}\boldsymbol{P}_{k,n}(\boldsymbol{\theta}_\phi)\boldsymbol{y}_n$ that appear in the vector $\boldsymbol{u}(\boldsymbol{\theta}_\phi, \hat{\sigma}_v^2(\boldsymbol{\theta}_\phi))$, or the sums $\sum_{n=0}^{N-1} \|\boldsymbol{y}_n\|^2$, and $\sum_{n=0}^{N-1} \|\boldsymbol{a}_n(\Delta_{k,n})\|^2$ appearing in (28).

ASML-3 is computationally most complex of all three solutions, because evaluating $\partial l(\boldsymbol{\theta}_\phi, \boldsymbol{\theta}_\sigma)/\partial \sigma_v^2$ requires one to recompute the matrix $\boldsymbol{R}_n(\boldsymbol{\theta}_\phi, \boldsymbol{\theta}_\sigma)$ in each iteration of steps 2-4 [c.f. (8)]. Our experience shows, however, that the number of iterations required for the convergence is small, which makes the algorithm's execution speed acceptable.

D. Unknown K case

In most situations the number of sources K is unknown. We consider the case of selecting K from multiple hypotheses, $K \in \{K_{\min}, \ldots, K_{\max}\}$, where $K_{\min}, K_{\max}$ are nonnegative integers satisfying $K_{\min} < K_{\max}$. The null hypothesis corresponds to $K = K_{\min}$. Denote by $\hat{\boldsymbol{\theta}}_{\phi|k}$, $k \in \{K_{\min}, \ldots, K_{\max}\}$, the DoA estimate/estimates obtained under the hypothesis that $K = k$, and by $\bar{l}(\hat{\boldsymbol{\theta}}_{\phi|k})$ the corresponding compressed log-likelihood. The proposed model order selection rule was adopted from [28, chap. 8.2], and has the advantage of having only one tunable parameter T

$$K_* = \underset{K \in \{K_{\min}, \ldots, K_{\max}\}}{\arg\max} \; 2\left[\bar{l}(\hat{\boldsymbol{\theta}}_{\phi|K}) - \bar{l}(\hat{\boldsymbol{\theta}}_{\phi|K_{\min}})\right] - p(K)T \;, \tag{30}$$

Outer search for DoAs
Maximize compressed log-likelihood

$$\hat{\boldsymbol{\theta}}_\phi = \arg\max_{\boldsymbol{\theta}_\phi} \bar{l}(\mathcal{Y}, \boldsymbol{\theta}_\phi)$$

Approximate evaluation of compressed log-likelihood
1. Set the initial value of noise variance $\hat{\sigma}_v^2$.
2. Solve

$$\boldsymbol{S}(\boldsymbol{\theta}_\phi)\hat{\boldsymbol{\sigma}}_s(\boldsymbol{\theta}_\phi) = \boldsymbol{u}(\boldsymbol{\theta}_\phi, \hat{\sigma_v}^2(\boldsymbol{\theta}_\phi))$$

where

$$[\boldsymbol{S}(\boldsymbol{\theta}_\phi)]_{ij} = \begin{cases} \sum_{n=0}^{N-1} w(\Delta_{i,n}) \boldsymbol{a}_n^{\mathrm{H}}(\Delta_{i,n}) \boldsymbol{a}_n(\Delta_{i,n}) & \text{for} \quad i = j \\ \sum_{n=0}^{N-1} w(\Delta_{i,n}) \frac{|\boldsymbol{a}_n^{\mathrm{H}}(\Delta_{i,n}) \boldsymbol{a}_n(\Delta_{j,n})|^2}{\boldsymbol{a}_n^{\mathrm{H}}(\Delta_{i,n}) \boldsymbol{a}_n(\Delta_{i,n})} & \text{for} \quad i \neq j \end{cases}$$

$$[\boldsymbol{u}(\boldsymbol{\theta}_\phi, \sigma_v^2)]_i = \sum_{n=0}^{N-1} w(\Delta_{i,n}) \left[\boldsymbol{y}_n^{\mathrm{H}} \boldsymbol{P}_{i,n}(\boldsymbol{\theta}_\phi) \boldsymbol{y}_n - \sigma_v^2 \right] .$$

If any element of $\hat{\boldsymbol{\sigma}}_s(\boldsymbol{\theta}_\phi)$ is negative, replace it with zero.
3. Set

$$\hat{\boldsymbol{\theta}}_\sigma(\boldsymbol{\theta}_\phi) \leftarrow [\hat{\boldsymbol{\sigma}}_s^{\mathrm{T}}(\boldsymbol{\theta}_\phi) \ \hat{\sigma}_v^2(\boldsymbol{\theta}_\phi)]^{\mathrm{T}}$$

4. Update noise variance estimate using one of the following:
(ASML-2)

$$\hat{\sigma}_v^2(\boldsymbol{\theta}_\phi) \leftarrow \frac{1}{NM} \left[\sum_{n=0}^{N-1} \|\boldsymbol{y}_n\|^2 - \sum_{k=1}^{K} \hat{\sigma}_{s,k}^2(\boldsymbol{\theta}_\phi) \sum_{n=0}^{N-1} \|\boldsymbol{a}_n(\Delta_{k,n})\|^2 \right]$$

(ASML-3)

$$\hat{\sigma}_v^2(\boldsymbol{\theta}_\phi) \leftarrow \hat{\sigma}_v^2(\boldsymbol{\theta}_\phi) + \alpha \sum_{n=0}^{N-1} \left[\boldsymbol{y}_n^{\mathrm{H}} \boldsymbol{R}_n^{-2}(\boldsymbol{\theta}_\phi, \hat{\boldsymbol{\theta}}_\sigma) \boldsymbol{y}_n - \mathrm{tr}\ \boldsymbol{R}_n^{-1}(\boldsymbol{\theta}_\phi, \hat{\boldsymbol{\theta}}_\sigma) \right]$$
$$\alpha = \hat{\sigma}_v^4(\boldsymbol{\theta}_\phi)/NM$$

If the updated estimate is below $\sigma_{v,\min}^2$, set

$$\hat{\sigma}_v^2(\boldsymbol{\theta}_\phi) \leftarrow \sigma_{v,\min}^2 .$$

5. Repeat steps 2-4 until convergence (i.e. until one step-changes of $\hat{\boldsymbol{\sigma}}_s(\boldsymbol{\theta}_\phi)$ and $\hat{\sigma}_v^2(\boldsymbol{\theta}_\phi)$ fall below a threshold) or until the maximum number of iterations is reached.
6. Compute

$$\bar{l}(\mathcal{Y}, \boldsymbol{\theta}_\phi) = -\left[\sum_{n=0}^{N-1} \log\det \boldsymbol{R}_n(\boldsymbol{\theta}_\phi, \hat{\boldsymbol{\theta}}_\sigma(\boldsymbol{\theta}_\phi)) + \boldsymbol{y}_n^{\mathrm{H}} \boldsymbol{R}_n^{-1}(\boldsymbol{\theta}_\phi, \hat{\boldsymbol{\theta}}_\sigma(\boldsymbol{\theta}_\phi)) \boldsymbol{y}_n \right]$$

Table I
SUMMARY OF ASML-2 AND ASML-3 ESTIMATORS

where $\bar{l}(\hat{\boldsymbol{\theta}}_{\phi|K})$ is the compressed maximized log-likelihood for the model with K sources and $p(K) = 2(K - K_{\min})$ is the difference in the number of parameters between the null hypothesis model and the K-source model.

The null hypothesis is rejected when any $2\left[\bar{l}(\hat{\boldsymbol{\theta}}_{\phi|K}) - \bar{l}(\hat{\boldsymbol{\theta}}_{\phi|K_{\min}})\right] - p(K)T$ is positive or, equivalently, when any

$$\frac{\bar{l}(\hat{\boldsymbol{\theta}}_{\phi|K}) - \bar{l}(\hat{\boldsymbol{\theta}}_{\phi|K_{\min}})}{K - K_{\min}} > T .$$

The order selection test statistic has the form

$$T_{\mathrm{OS}} = \max_{K \in \{K_{\min}, \ldots, K_{\max}\}} \frac{\bar{l}(\hat{\boldsymbol{\theta}}_{\phi|K}) - \bar{l}(\hat{\boldsymbol{\theta}}_{\phi|K_{\min}})}{K - K_{\min}}$$

and has the limiting ($N \to \infty$) distribution equal to that of [28, chap. 8.2]

$$\bar{T}_{\mathrm{OS},r} = \max_{m=2,4,\ldots,2r} (Z_1^2 + Z_2^2 + \cdots + Z_m^2)/m \qquad (31)$$

where $r = K_{\max} - K_{\min}$ and $Z_1, Z_2, \ldots, Z_m$ are independent standard normal variables.

Note that for $r = 1$ (e.g., $K_{\min} = 1$, $K_{\max} = 2$), the rule (30) corresponds to the classical generalized likelihood ratio test, and the test statistic is, asymptotically, χ_2^2 distributed [25], [31]. Therefore, in this case one may adjust the threshold level on the basis of the χ^2 distribution tables and desired false alarm rates. The remaining distributions for $r > 1$ can be obtained using straightforward Monte Carlo simulations – see Fig. 1 for the cases $r = 1, 2, 3, 4$. Observe that for sufficiently high T there is practically no difference between all the plots, and one obtains that setting $T = 4.6$ should ensure $P_{FA} \approx 0.01$, while $T = 6.9$ should results in $P_{FA} \approx 0.001$. However, one should treat these results with caution. First, the asymptotic distribution may be inaccurate (even for $N \to \infty$) because the source variances are at the boundary of their support. Second, we intend to perform the test (30) using the compressed likelihood obtained from the proposed ASML estimators, which may result in some additional mismatch. To verify characteristics of the proposed test, we ran additional Monte Carlo simulations, results of which will be reported in the next section.

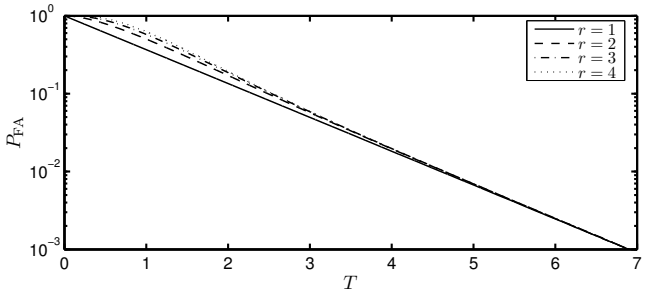

Figure 1. Limiting false alarm rates as functions of threshold T for the multiple hypothesis tests with different $r = K_{\max} - K_{\min}$.

Remark: Similar to the nonrotating array case [18], if one assumes a dense grid of (excessinve in number) hypothetical source DoAs, the exact maximum likelihood approach (4) achieves a sparse a solution in the sense that most source variances take values close to zero, and the nonzero ones are concentrated around the true source DoA. While this property could potentially be used to estimate the number of sources, we point out that similar behavior was not observed for the ASML estimators and that the computational complexity of the exact maximum likelihood approach is too high to make it practical. Therefore, more research in this direction is needed.

V. RESULTS OF MONTE CARLO SIMULATIONS

A. Influence of source correlation on exact maximum likelihood method

To demonstrate that the analytical approach of investigating the influence of source correlation presented in Section II.C is valid, and to demonstrate that the maximum likelihood estimator based on the model of uncorrelated sources (1) suffers only small losses for low and moderate levels of source correlation, we arranged a straightforward Monte Carlo simulation.

Following [21], we simulated a system that employs a standard uniform linear array (half-wavelength element spacing) consisting of $L = 64$ elements, and synthesizes the sum and difference beampatterns using the rectangular (boxcar)

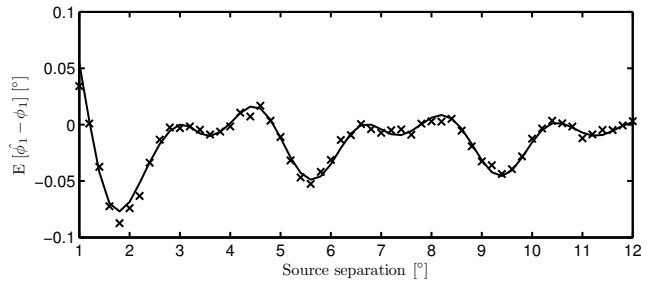

Figure 2. Comparison of analytic predictions of the narrow estimator bias (solid line) with its actual values (crosses) for several values of source separation.

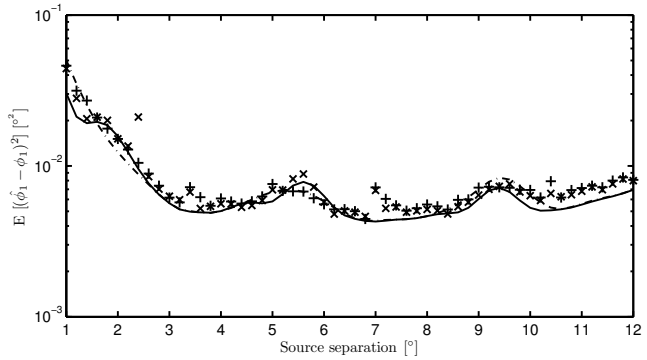

Figure 3. Comparison of analytic predictions of mean-squared estimation errors of the narrow and wide estimator obtained using (15)-(16) (solid line and dash-dotted line, respectively) with their actual values (crosses and pluses, respectively) for several values of source separation.

taper. The resultant array steering vector $\boldsymbol{a}(\Delta_\phi)$, which reads

$$\boldsymbol{a}(\Delta\phi) \;=\; \frac{1}{L}\left[\begin{array}{cc} \dfrac{F_L(\Delta\phi)-1}{F(\Delta\phi)-1} & -j\dfrac{\left(F_{L/2}(\Delta\phi)-1\right)^2}{F(\Delta\phi)-1} \end{array}\right]^{\mathrm{T}},$$

where $F_L(\Delta\phi) = e^{jL\pi\sin(\Delta\phi)}$ and $F(\Delta\phi) = F_1(\Delta\phi)$, has the 3-dB beamwidth of $1.6°$.

The simulated rotation of the array covered azimuth angles between $-6.2°$ and $+6.2°$, during which $N = 32$ simulated observations were generated at a grid of equally-spaced array positions.

We simulated two moderately correlated ($\rho = 0.5e^{j\pi/6}$) sources with signal to noise ratios $\sigma^2_{s,1}/\sigma^2_v = \sigma^2_{s,2}/\sigma^2_v$ equal to 13 dB. The sources were placed symmetrically about the zero azimuth, $\phi_1 = -\phi_2$, and their angular separation was varied from $1°$ to $12°$ in $0.1°$ steps. For each separation we ran 600 independent simulations.

We compared bias and the mean-squared estimation error of the narrow and wide estimators against each other and with their analytical predictions given in (15)-(16). Both estimators were implemented in their exact forms, i.e., without using any approximations (see Section V.D for results concerning the ASML estimators). The results, in the form of plots, are shown in Fig. 2-3.

Observe that the theoretical analysis and the simulations show good agreement, particularly in terms of the bias of the narrow estimator. We verified that the 95% confidence interval for the Monte Carlo estimate of the bias covers the theoretical value for all simulated separations except at $1.8°$. The discrepancy between the MSE values predicted by the theory and the actual results can be attributed to somewhat too small number of observations. Since the beamwidth is $1.6°$ and observations are taken every $0.4°$ there is only a handful of data from the mainlobe of the beampattern, which results in both estimators exceeding their theoretical variance by a small factor – see the next subsection for more results concerning this matter.

Furthermore, note that the simpler narrow estimator compares quite favorably against the wide one in terms of the mean-squared error in the entire range of simulated source separations. Taking into account that the narrow estimator is computationally considerably less complex than the wide estimator, there is hardly any reason to employ the latter, unless one expects higher levels of source correlation – see Section V.E for additional results and discussion of this issue.

B. Accuracy of ASML estimators

We compared the accuracy, in the sense of the mean-squared angle estimation error, of ASML-2 and ASML-3 estimators with the Crámer-Rao lower bound (CRLB) for the two-source case, $K = 2$. The performance of the ASML-1 estimator was not tested for two reasons: First, the requirement that $M > K$ is not satisfied in the monopulse configuration. Second, in [30] we already demonstrated that in the single-source case it is comparable with the exact maximum likelihood approach.

In both cases, the optimization of the compressed log-likelihood consisted of two stages: preliminary search over a rather coarse grid of angles ($0.5°$ steps), followed by the final search using the Nelder-Mead simplex method. We also tested a mixed solution, which we call ASML-2+3, which combines ASML-2 in the preliminary search and ASML-3 in the final search stages.

All algorithms employed the same weight function $w(\Delta_{i,n}) = \|\boldsymbol{a}_n(\Delta_{i,n})\|^2$. Although we make no claim that this choice is optimal, it works much better than the straightforward uniform weighting, in which case we faced rather severe problems with the stability of the estimates of the source and noise variances at certain combinations of angles. Another advantage of this choice is that it allows one to eliminate any risk of division by zero occurring in the computation of the matrix $\boldsymbol{S}(\boldsymbol{\theta}_\phi)$.

We considered three scenarios, where we varied only one factor at each time: the signal to noise ratio $\mathrm{SNR} = \sigma^2_{s,1}/\sigma^2_v = \sigma^2_{s,2}/\sigma^2_v$, the number of observations N, and the source separation. In all cases the we simulated $K = 2$ independent sources and the antenna rotation covered the angles between $-4°$ and $+4°$. The "default" values of N, SNR, and source separation are 48, 10 dB, and $2°$ ($\phi_1 = -0.5°$, $\phi_2 = 1.5°$), respectively. The results are shown in Fig. 4-6.

Observe that ASML-2 performs uniformly worse than ASML-3, and that its accuracy depends more on the number of observations than on the SNR. ASML-3, on the other hand, nearly reaches the CRLB. Finally, the combination of ASML-2 and ASML-3 is almost equivalent to ASML-3 (with the exception of slightly worse behavior in the threshold area). Since the computational complexity of the mixed approach is

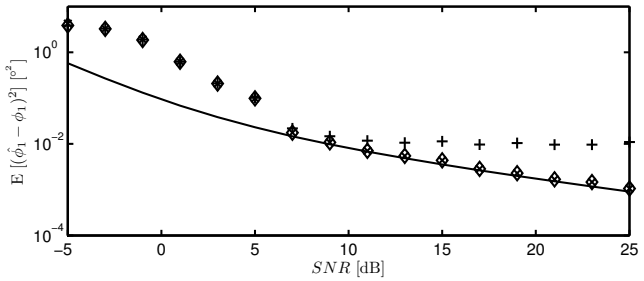

Figure 4. Comparison of mean-squared angle estimation errors of ASML-2 (pluses), ASML-3 (diamonds), and ASML-2+3 (crosses) with Crámer-Rao lower bound for different signal to noise ratios.

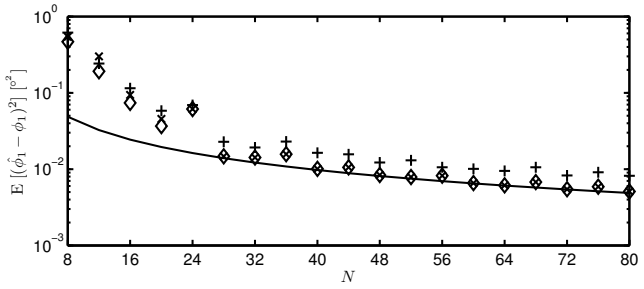

Figure 5. Comparison of mean-squared angle estimation errors of ASML-2 (pluses), ASML-3 (diamonds), and ASML-2+3 (crosses) with Crámer-Rao lower bound for different number of observations.

considerably lower, we recommend this option for practical applications.

C. Detection and resolution of multiple sources using ASML estimators

Fig. 7 shows the comparison of the limiting distribution (31) for the multiple-hypothesis test $K \in \{K_{\min}, \ldots, K_{\max}\}$ ($K_{\min} = 0$, $K_{\max} = 3$) with the Monte Carlo estimates (50000 independent simulations) of the false-alarm probability yielded by the ASML-3 estimator under antenna rotation from $-5°$ to $+5°$, and $N \in \{8, 32, 64, 128\}$. The results confirm that the limiting distribution (31) is reasonably accurate for wide range of N – to achieve a desired false alarm rate one should increase T by about 0.6 from the analytic value. Similar results were observed for different values of r, with the discrepancy getting slightly smaller as r decreased.

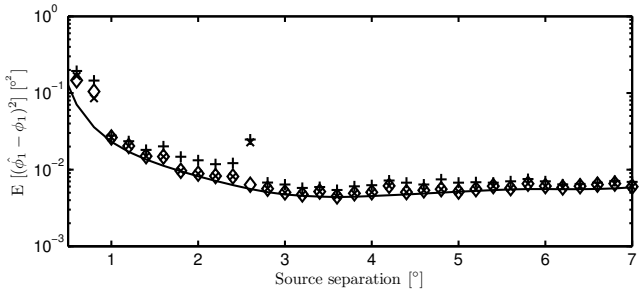

Figure 6. Comparison of mean-squared angle estimation errors of ASML-2 (pluses), ASML-3 (diamonds), and ASML-2+3 (crosses) with Crámer-Rao lower bound for different source separations.

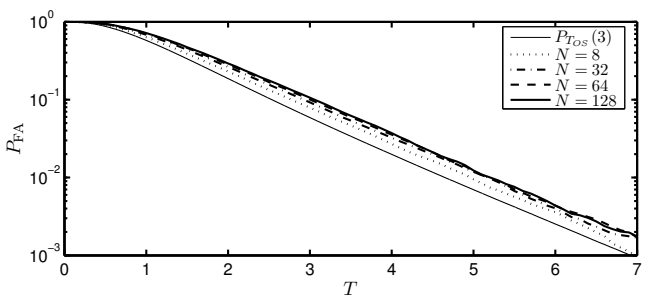

Figure 7. Comparison of asymptotic probability of false alarm $P_{T_{OS}}(3)$ for $K \in \{0, 1, 2, 3\}$ (thin solid line) and actual results for ASML-3 estimator and several values of N (remaining lines) as functions of threshold T.

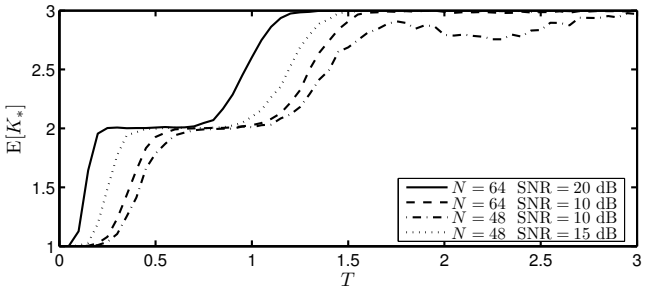

Figure 8. Average number of detected sources as function of source separation for ASML-3 estimator, three sources, $K \in \{0, 1, 2, 3\}$, $T = 6.9$ and several values of N and SNR

Fig. 8 shows the average number of sources detected for the case of three equally strong sources placed symmetrically about 0, $\phi_2 - \phi_1 = \phi_3 - \phi_2$, $\phi_2 = 0$ as a function of the source separation for several choices of SNR and N (each point was obtained from 1000 independent simulations, $T = 6.9$). Generally, one can observe reasonable behavior of the combination of the estimator and the proposed decision rule, with the number of detected sources gradually growing from 1 to its actual value as the source separation grows. Interestingly enough, for small N or SNR and source separation taking such values that the observation variances are close to constant during entire array rotation, the number of sources declared by (30) could drop to zero. This unexpected behavior occurs because under such adverse conditions there is not enough evidence to discriminate this case from pure noise reliably.

D. Influence of source correlation on ASML estimators

An important property of the ASML-2 and -3 estimators is that, similar to the exact maximum likelihood approach, they exhibit a gradual degradation, rather than breakdown, when sources are correlated. Fig. 9 compares the analytic predictions of MSE given by (15) with the the corresponding Crámer-Rao lower bound [obtained by assuming that ρ is included into the vector of parameters; recall that the formula (16) is obtained by including ρ in the vector of parameters and setting it to 0] as functions of $|\rho|$, for $\arg \rho = \pi/6$, $N = 64$, antenna rotation from $-6.2°$ to $6.2°$, two sources at $\pm 1°$, each with SNR equal to 10 dB. Observe that, according to (15), the narrow estimator should offer best performance until $|\rho|$ reaches 0.25, and that

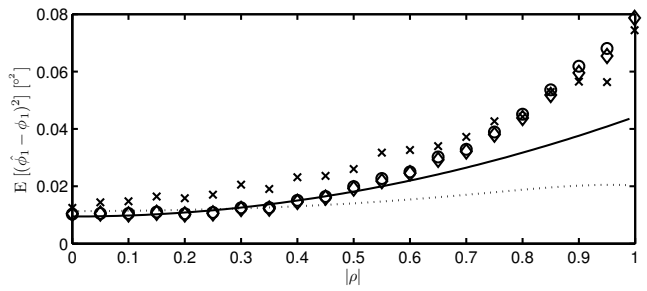

Figure 9. Comparison of analytic predictions of mean-squared estimation errors of the narrow obtained using (15) (solid line) with the exact Crámer-Rao lower bound for the wide model (dotted line) and actual mean-squared errors of exact maximum likelihood estimator, ASML-2 and ASML-3 estimator (diamonds, crosses, and circles, respectively).

for $|\rho| = 1$ the formula (15) predicts that the MSE of the narrow estimator will reach about 1.6 times the CRLB.

The same figure also includes the Monte Carlo estimates of the MSE obtained for the exact maximum likelihood estimator, as well as for the proposed ASML-2 and ASML-3 estimators (500 independent simulations were conducted for each value of $|\rho|$). Observe that the accuracy of ASML-3 is very close to the exact maximum likelihood approach, and that the MSE values for these two algorithms agree well with analytic predictions up to $|\rho| = 0.5$. For higher levels of source correlation the rise in the MSE is sharper than predicted by (15), but the estimator breakdown does not occur, and the MSE eventually reaches about four times the CRLB.

Remark: Fig. 9 can be regarded as a pessimistic assessment of the estimators's performance at high source correlations, because the bias caused by dropping ρ from the model depends not only on $|\rho|$, but also on $\arg\rho$ [c.f. (13)] – in the situation presented the worst-case value of $\arg\rho$ is approximately $\pi/12$ and the resultant bias is only 2% higher than at $\arg\rho = \pi/6$.

E. Comparison with existing solutions

We compared the ASML-2+3 estimator with the AML and AML-RELAX estimators from [21]. The AML and AML-RELAX estimators were were derived using the deterministic maximum likelihood approach specifically for coherent radar systems. Assuming K targets, the AML estimator performs $2K$-dimensional search in the space of DoAs and target Doppler frequencies. The AML-RELAX estimator is based on the assumption that targets are sufficiently well separated in Doppler and employs K two-dimensional searches. Clearly, if the data-generation mechanism matches that assumed by the AML and AML-RELAX estimators, the ASML-2+3 estimator is inferior in every aspect because it neglects the Doppler information. Fig. 10 shows the comparison of the ASML-2+3, AML, and AML-RELAX under two another data-generation mechanisms. Observe that if targets have the same Doppler the proposed solution is quite superior to AML-RELAX at some source separations. Moreover, if we use our data-generation mechanism, both AML estimators are simply fail, because of their reliance on the presence of Doppler in the observations.

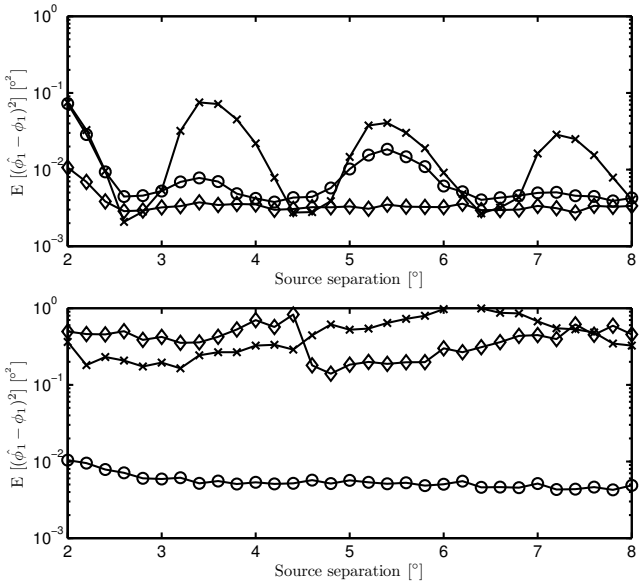

Figure 10. Comparison of mean-squared errors yielded by AML (diamonds), AML-RELAX (crosses) and ASML-2+3 (circles) estimators under data generated using deterministic model with same Doppler (top plot) and stochastic model (bottom plot).

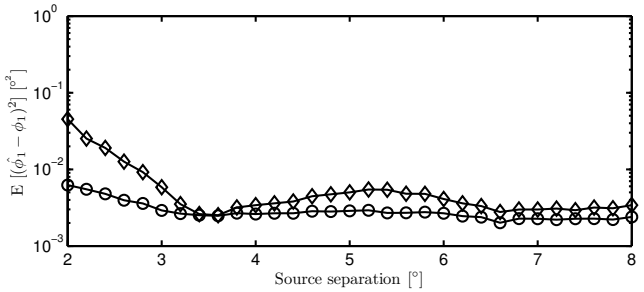

Figure 11. Comparison of mean-squared errors yielded by Capon-like estimator from [23] (diamonds) and ASML-2+3 (circles) under stochastic model.

Fig. 11 shows the comparison of accuracy of the ASML-2+3 estimator with the Capon-like estimator for rotating arrays proposed in [23]. The comparison was performed using data generated from the model (1). Even though the proposed approach performs uniformly better, the most important differences are, perhaps, of qualitative nature. The Capon-like estimator requires a parametric model of the array manifold $\boldsymbol{a}(\Delta\phi)$ in the form of a polynomial (we used a polynomial with $K = 8$ terms), and the size of the polynomial dictates the minimum number of observations ($N \geq MP$). Moreover, if K is too large, the implementation of the estimator becomes challenging because special attention to the numerical accuracy is needed. Such problems simply do not exist with the ASML-3 estimator, which can even use a nonparametric array manifold model, such as a lookup table. Moreover, as demonstrated, the ASML-3 estimator can work with fully coherent sources and – last but not least – supports the frequency-agile mode of operation.

VI. Conclusions

We presented new results that concern DoA estimation for rotating arrays. The proposed approach is based on the stochastic maximum likelihood framework and employs a model that assumes the sources are incoherent. We showed that under mild conditions this approach is comparable, in terms of accuracy, with the more complex estimator that includes the source correlation coefficient. We also showed that the simpler estimator is fairly insensitive to small errors in estimating the source and noise variances, which allowed us to propose its three fast approximations. The case of unknown number of sources was addressed with an extension of the generalized likelihood ratio approach. Behavior of all solutions was verified using Monte Carlo simulations, which pointed to the combination of the ASML-2 and ASML-3 algorithms as the best option.

References

[1] D. K. Barton, *Radar System Analysis and Modeling*. Artech House, Inc., 2005.

[2] M. Skolnik, *Introduction to Radar Systems*. McGraw-Hill, 2002.

[3] W.-D. Wirth, *Radar Techniques Using Array Antennas*, ser. Radar, Sonar & Navigation. Institution of Engineering and Technology, 2013.

[4] J. Capon, "High resolution frequency-wavenumber spectrum analysis," *Proceeding of the IEEE*, vol. 57, p. 1408–1418, 1969.

[5] J. Li, P. Stoica, and Z. Wang, "On robust Capon beamforming and diagonal loading," *IEEE Transactions on Signal Processing*, vol. 51, no. 7, pp. 1702–1715, 2003.

[6] R. Schimidt, "Multiple emiter location and signal parameter estimation," *IEEE Transactions on Antennas and Propagation*, vol. 34, no. 3, pp. 276–280, 1986.

[7] B. D. Rao and K. V. S. Hari, "Performance analysis of root-music," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, no. 12, pp. 1939–1949, 1989.

[8] R. Roy and T. Kailath, "ESPRIT – Estimation of Signal Parameters via Rotational Invariance Techniques," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 37, no. 7, pp. 984–995, 1989.

[9] V. V. Reddy, M. Mubeen, and B. P. Ng, "Reduced-complexity super-resolution DOA estimation with unknown number of sources," *IEEE Signal Processing Letters*, vol. 22, no. 6, pp. 772–776, 2015.

[10] T. Yardibi, J. Li, P. Stoica, M. Xue, and A. B. Baggeroer, "Source localization and sensing: a nonparametric iterative adaptive approach based on weighted least squares," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 46, pp. 425—-443, 2010.

[11] H. L. van Trees, K. L. Bell, and Z. Tian, *Detection, Estimation and Modulation Theory, Part IV: Optimum Array Processing*. John Wiley & Sons, 2002.

[12] O. Besson and P. Stoica, "Decoupled estimation of doa and angular spread for a spatially distributed source," *IEEE Transactions on Signal Processing*, vol. 48, no. 7, pp. 1872–1882, 2000.

[13] O. Besson, F. Vincent, P. Stoica, and A. B. Gershman, "Approximate maximum likelihood estimators for array processing in multiplicative noise environments," *IEEE Transactions on Signal Processing*, vol. 48, no. 9, pp. 2506–2518, 2000.

[14] F. Gini and F. Lombardini, "Multibaseline cross-track sar interferometry: a signal processing perspective," *IEEE Aerospace and Electronic Systems Magazine*, vol. 20, no. 8, pp. 71–93, 2005.

[15] F. Gini, F. Lombardini, and M. Montanari, "Layover solution in multibaseline sar interferometry," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 38, no. 4, pp. 1344–1356, 2002.

[16] Y. D. Zhang, M. G. Amin, and B. Himed, "Sparsity-based DOA estimation using co-prime arrays," in *Proc 2013 International Conference of Acoustics, Speech and Signal Processing*, 2013, pp. 3967–3971.

[17] S. Qin, Y. D. Zhang, and M. G. Amin, "Generalized coprime array configurations for direction-of-arrival estimation," *IEEE Transactions on Signal Processing*, vol. 63, no. 6, pp. 1377–1390, 2015.

[18] Z. Yang, J. Li, P. Stoica, and L. Xie, "Chapter 11 - sparse methods for direction-of-arrival estimation," in *Academic Press Library in Signal Processing, Volume 7*, R. Chellappa and S. Theodoridis, Eds. Academic Press, 2018, pp. 509 – 581.

[19] A. Farina, F. Gini, and M. Greco, "DOA estimation by exploiting the amplitude modulation induced by antenna scanning," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 38, no. 4, pp. 1276–1286, 2002.

[20] F. Gini, M. Greco, and A. Farina, "Multiple radar targets estimation by exploiting induced amplitude modulation," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 39, no. 4, pp. 1316–1321, 2003.

[21] M. Greco, F. Gini, and A. Farina, "Joint use of sum and delta channels for multiple radar target doa estimation," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 43, no. 3, pp. 1146–1154, 2007.

[22] M. Greco, F. Gini, A. Farina, and L. Timmoneri, "Radar target DOA estimation: Moving window vs AML estimator," in *Proc. 2008 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2008)*, 2008, pp. 1501–1504.

[23] M. Meller and K. Stawiarski, "Capon-like DoA estimator for rotating arrays," in *Proc. 2019. International Radar Conference*, Toulon, France, 2019.

[24] A. Zeira and B. Friedlander, "Direction finding with time-varying arrays," *IEEE Transactions on Signal Processing*, vol. 43, no. 4, pp. 927–937, 1995.

[25] H. L. van Trees, K. L. Bell, and Z. Tian, *Detection, Estimation and Modulation Theory, Part I: Detection, Estimation, and Filtering Theory*. John Wiley & Sons, 2013.

[26] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

[27] A. G. Jaffer, "Maximum likelihood direction finding of stochastic sources: a separable solution," in *Proc. 1988 International Conference on Acoustics, Speech, and Signal Processing (ICASSP-88)*, 1988, pp. 2893–2896.

[28] G. Claeskens and N. L. Hjort, *Model Selection and Model Averaging*, ser. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 2008.

[29] ——, "The focused information criterion," *Journal of the American Statistical Association*, vol. 98, pp. 900–916, 2003.

[30] M. Meller, K. Stawiarski, and B. Pikacz, "Azimuth estimator for a rotating array radar with wide beam," *Proc. SPIE*, vol. 10715, pp. 10 715 – 10 715 – 5, 2018. [Online]. Available: https://doi.org/10.1117/12.2317865

[31] S. S. Wilks, "The large-sample distribution of the likelihood ratio for testing composite hypotheses," *Annals of Mathematical Statistics*, no. 1, 1938.