



HAL
open science

Joint Source Estimation and Localization

Souleymen Sahnoun, Pierre Comon

► **To cite this version:**

Souleymen Sahnoun, Pierre Comon. Joint Source Estimation and Localization. 2014. hal-01005352v1

HAL Id: hal-01005352

<https://hal.science/hal-01005352v1>

Preprint submitted on 12 Jun 2014 (v1), last revised 20 Feb 2015 (v2)

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.

Joint Source Estimation and Localization

Souleymen Sahnoun, *Member, IEEE*, and Pierre Comon, *Fellow, IEEE*

Abstract

The estimation of directions of arrival is formulated as the decomposition of a 3-way array into a sum of rank-one terms, which is possible when the receive array enjoys some geometrical structure. The main advantage is that this decomposition is essentially unique under mild assumptions, if computed exactly. The drawback is that a low-rank approximation does not always exist. Therefore, a constraint is first introduced that ensures the existence of the latter best approximate. Then Cramér-Rao bounds are derived for localization parameters and source signals, assuming the others are nuisance parameters; some inaccuracies found in the literature are pointed out. Performances are eventually compared with reference algorithms such as ESPRIT, in the presence of additive Gaussian noise, with possibly non circular distribution.

Index Terms

multi-way array, tensor decomposition, source localization, antenna array processing, low-rank approximation, complex Cramer-Rao bounds

I. INTRODUCTION

Estimation of Directions of Arrival (DoA) is a central problem in antenna array processing, including in particular radar, sonar, or telecommunications [1]. Over the last decades, several DoA estimation tools have been developed, ranging from nonparametric Fourier-based methods to parametric high-resolution techniques. The latter techniques, including linear prediction-based methods and subspace methods, are often preferred to nonparametric ones since they achieve high resolution estimates. Recently, methods based on sparse approximations have been proposed, which are considered as semi-parametric [2], [3].

Traditional subspace approaches such as MUSIC (multiple signal classification) are based on low-rank approximation of the covariance matrix of observations, and on detecting points of minimal distance with

Work funded by the FP7 programme 2007–2013, ERC Grant Agreement no. 320594, DECODA project. Partly presented during the 2014 SAM Workshop.

S. Sahnoun and P. Comon are with the Gipsa-Lab, 11 rue des Mathématiques, Domaine Universitaire BP 46, 38402 Saint Martin d'Hères Cedex, France (e-mail: firstname.name@gipsa-lab.grenoble-inp.fr)

Manuscript received June 12, 2014.

the so-called array manifold [4] [5]. These approaches hence assume that (i) the measurements are weakly stationary over sufficiently long observation lengths, (ii) the number of sources of interest is smaller than the number of sensors, and (iii) spatial responses of all sensors are known, and in particular their location (in other words, the sensor array needs to be calibrated). Another well-known subspace method is the ESPRIT (estimation of signal parameters via rotational invariance techniques) algorithm [6] which is applicable to sensor arrays consisting of two identical displaced subarrays. The displacement vector between the two subarrays should be known, whereas, unlike MUSIC-type algorithms, the calibration information of each subarray is not required.

In [7], a deterministic approach has been proposed, which permits not only to work with short data lengths (and hence less stationary sources), but also to localize more sources than sensors present in the reference array. This approach is based on the same rotational invariance as exploited in ESPRIT [6], but can handle more than one displacement. It consists in storing the measurements in a 3-way array, and to decompose it into a sum of rank-one terms. One very interesting by-product of [7] is that source copies are also delivered for free, without any further estimation stage.

The most popular model for fitting a 3-way array using a sum of rank-one terms is the canonical polyadic (CP) decomposition. Contrary to the decomposition of matrices in rank-one terms, the CP decomposition is essentially unique under mild conditions. However, a best-fitting CP decomposition may not exist for some tensors [8]. In this case, trying to find a low-rank tensor approximation results in diverging components.

In this paper we revisit the approach of [7], where the important issue of the existence of a low-rank tensor approximation has been neglected, so that the latter approach is actually ill posed. This fact has been already pointed out in [8], and additional constraints have been suggested, which enjoy a reliable physical meaning and at the same time ensure existence of a solution. In the present work, we shed some light on conditioning and algorithmic issues. Our main contributions are the following. We first propose a new differentiable constraint which guarantees existence of the low-rank tensor approximation. Then, we derive the expressions of the Cramér-Rao bounds (CRB) related to the localization problem in the presence of nuisance parameters estimated by CP decomposition methods. Note that the CRB for CP decomposition of complex-valued and real-valued tensors were respectively studied in [9] and [10], but without assuming that one of matrix factors is parameterized by angles of arrival.

The remainder of this paper is organized as follows. In the next section, we introduce notation and formulate the DoA estimation as a low-rank decomposition of a 3-way array. Then we define in Section III the constraint ensuring existence of the best approximation. An optimization algorithm is described

in Section IV, and Cramér-Rao bounds of DoA parameters are derived in Section V. Finally, we show the benefit of the proposed algorithm via numerical experiments, and compare results to Cramér-Rao bounds.

II. MODELING AND NOTATION

In the following, vectors will be denoted by bold lowercases, *e.g.* \mathbf{a} , whereas matrices or higher-order arrays will be denoted by bold uppercases, *e.g.* \mathbf{A} . Moreover, \mathbf{a}_r will denote the r th column of matrix \mathbf{A} .

Suppose that R narrow-band radiating sources impinge on an array of sensors, formed of L subarrays of K sensors each. We make the far-field assumption, that is, we assume that sources are located sufficiently far from the array, compared to the array dimensions, so that waves can be considered as plane. The key assumption made in [6], [7], [8] is that, taking one subarray as reference, every subarray can be deduced from the reference one by an unknown translation in space, defined by some vector δ_ℓ of \mathbb{R}^3 , $1 \leq \ell \leq L$, $\delta_1 \stackrel{\text{def}}{=} \mathbf{0}$.

Denote $\varsigma_r(t)$ the signal transmitted by the r th source, \mathbf{d}_r its DoA viewed from the array, and $s(\mathbf{y}, t)$ the signal measured at a point in space defined by its coordinates \mathbf{y} (we consider complex envelopes about the central frequency). Then we have:

$$s(\mathbf{y}, t) = \sum_{r=1}^R \varsigma_r(t) a_r(\mathbf{y}), \quad a_r(\mathbf{y}) \stackrel{\text{def}}{=} \exp\{j \frac{\omega}{C} \mathbf{y}^\top \mathbf{d}_r\} \quad (1)$$

where ω is the central pulsation of the narrow-band waves, C the wave celerity, and $j = \sqrt{-1}$. Because waves are plane and narrow-band, the signal measured at another point $\mathbf{y} + \boldsymbol{\tau}$, deduced from \mathbf{y} by a translation $\boldsymbol{\tau}$ takes the form:

$$s(\mathbf{y}, \boldsymbol{\tau}, t) = \sum_{r=1}^R a_r(\mathbf{y}) b_r(\boldsymbol{\tau}) \varsigma_r(t), \quad b_r(\boldsymbol{\tau}) \stackrel{\text{def}}{=} \exp\{j \frac{\omega}{C} \boldsymbol{\tau}^\top \mathbf{d}_r\} \quad (2)$$

In other words, function $s(\mathbf{y}, \boldsymbol{\tau}, t)$ decomposes into a sum of R simpler functions whose variables separate.

Now, if we discretize the \mathbb{R}^3 space with the above defined subarrays, and take M time samples, we end up with a multi-linear relationship¹ in finite dimensional spaces. In fact, let \mathbf{p}_k be the coordinate vector of the k th sensor of the reference subarray, and δ_ℓ the translation defining the location of the ℓ th subarray, $1 < \ell \leq L$. Then signal (2) can be stored in a $K \times L \times M$ three-way array, which follows the model below, possibly corrupted by additive noise:

$$T_{k\ell m} = \sum_{r=1}^R \lambda_r A_{kr} B_{\ell r} S_{mr} \quad (3)$$

¹See *e.g.* [11] for a definition of multi-linearity.

where $A_{kr} = \frac{1}{\sqrt{K}} \exp(j\frac{\omega}{c} \mathbf{p}_k^T \mathbf{d}_r)$, $B_{\ell r} = \frac{1}{\sqrt{L}} \exp(j\frac{\omega}{c} \delta_\ell^T \mathbf{d}_r)$, $S_{mr} = \varsigma_r(t_m) / \|\varsigma_r\|$, and $\lambda_r = \sqrt{KL} \|\varsigma_r\|$. Note that \mathbf{a}_r , \mathbf{b}_r and \mathbf{s}_r are hence unit L^2 -norm vectors.

Model (3) is related to the Canonical Polyadic decomposition (CP)², which consists of decomposing a tensor \mathbf{T} into a sum of *decomposable* tensors. For the sake of convenience, equation (3) is rewritten in vector form as

$$\mathbf{t} = \sum_{r=1}^R \lambda_r \mathbf{a}_r \boxtimes \mathbf{b}_r \boxtimes \mathbf{s}_r \quad (4)$$

where \boxtimes denotes the Kronecker³ product, and $\mathbf{t} = \text{vec}\{\mathbf{T}\}$ is a column vector of dimension KLM containing the entries of the 3-way array \mathbf{T} .

III. EXISTENCE AND UNIQUENESS

The goal is to identify the directions of arrival (DoA), ψ_r , of the R impinging plane waves and to estimate corresponding transmitted source signals $\varsigma_r(t_m)$ up to a scaling factor, given the whole array \mathbf{T} . Clearly, a sufficient condition is to be able to identify all parameters in the RHS of (4).

A. Low rank approximation

Actually, observations are corrupted by noise, so that (3-4) do not hold exactly. A natural idea is then to fit model (4) by minimizing the error

$$\Upsilon(\mathbf{A}, \mathbf{B}, \mathbf{S}; \mathbf{\Lambda}) = \left\| \mathbf{t} - \sum_{r=1}^R \lambda_r \mathbf{a}_r \boxtimes \mathbf{b}_r \boxtimes \mathbf{s}_r \right\|^2 \quad (5)$$

where $\mathbf{\Lambda}$ denotes the diagonal matrix containing the λ_r 's, and $\|\cdot\|$ the L^2 -norm. Minimizing error (5) means finding the best rank- R approximate of \mathbf{T} and its CP decomposition. However, the infimum of Υ may not be reached; see *e.g.* [13], [8] and references therein. The reason is that the set of rank- R tensors is not closed for $R > 1$. The idea we promote here is to impose an additional constraint that will ensure the existence of a minimum, as elaborated in the next section. In addition, from the physical point of view, one can make the following observations:

- sources that are totally correlated need to be localized separately only if they are sufficiently well angularly separated. In that case they correspond to multi-paths of the same radiating source.
- sources that are located in the same direction need to be estimated separately only if they are sufficiently decorrelated. In the latter case, they correspond to different sources.

²also sometimes called Candecomp/Parafac in Psychometry or Chemometrics [12].

³This notation is chosen to make the distinction between tensor and Kronecker products, which are sometimes mixed up [11].

- otherwise, one can assimilate highly correlated sources arriving from close directions to a single *fat source*, spread out in space.

The purpose of the section is to formalize these constraints.

B. Coherences

As in the compressed sensing literature [14], [15], we define the *coherence of a set* of unit norm vectors as the largest value of cross scalar products:

$$\mu_A = \sup_{k \neq \ell} |\mathbf{a}_k^H \mathbf{a}_\ell| \quad (6)$$

The coherence of matrix \mathbf{A} is defined this way, if \mathbf{a}_k denote its (unit norm) columns. Coherences of matrices \mathbf{B} and \mathbf{S} are defined similarly, and denoted by μ_B and μ_S , respectively. Let \mathbf{G} be the $R \times R$ Gram matrix defined by:

$$G_{pq} = (\mathbf{a}_p^H \mathbf{a}_q)(\mathbf{b}_p^H \mathbf{b}_q)(\mathbf{s}_p^H \mathbf{s}_q)$$

Then for given matrices \mathbf{A} , \mathbf{B} and \mathbf{S} , the optimal value Λ^o minimizing error Υ is obtained by cancelling the gradient of (5) w.r.t. Λ , which leads to the linear system:

$$\mathbf{G} \boldsymbol{\lambda}^o = \mathbf{f}, \quad (7)$$

where $\boldsymbol{\lambda}^o = \text{diag}(\Lambda^o)$ and vector \mathbf{f} in the right hand side is defined by the contraction $f_r = \sum_{ijk} T_{ijk} A_{ir}^* B_{jr}^* S_{kr}^*$, $1 \leq r \leq R$. Equation (7) shows that coherences play a role in the conditioning of the minimization problem. Also note that only the *product* between coherences appears, and not coherences individually.

C. Existence

We are now in a position to state conditions of existence. It has been shown in [8] that if

$$\mu_A \mu_B \mu_S < \frac{1}{R-1} \quad (8)$$

then the infimum of (5) is reached. This happens because error (5) becomes coercive as soon as (8) is satisfied. And it must then reach its minimum since it is continuous.

Constraint (8) needs some care because it involves max operators, which are not differentiable. For this reason, we propose to use the fact that the L^∞ norm can be bounded by $L^{2\rho}$ norms, and approximated for large values of ρ :

$$\|\mathbf{z}\|_\infty = \max_k \{z_k\} \leq \|\mathbf{z}\|_{2\rho} = \left(\sum_k z_k^{2\rho} \right)^{1/2\rho}, \quad \forall \rho \geq 1$$

for $z_k \in \mathbb{R}^+$. Applying this inequality to $z_k \equiv |\mathbf{a}_p^H \mathbf{a}_q|$ allows to bound coherences above by a differentiable quantity, so that another (somewhat more constraining) sufficient condition can be obtained. More precisely:

$$\mu_A \leq \mu(\mathbf{A}, \rho) \stackrel{\text{def}}{=} \left(\sum_{p < q} |\mathbf{a}_p^H \mathbf{a}_q|^{2\rho} \right)^{1/2\rho}$$

We subsequently call \mathcal{C}_ρ the constraint obtained by replacing the max operators by the $L^{2\rho}$ norms in constraint (8):

$$\mathcal{C}_\rho \stackrel{\text{def}}{=} 1 - R + \mu(\mathbf{A}, \rho)^{-1} \mu(\mathbf{B}, \rho)^{-1} \mu(\mathbf{S}, \rho)^{-1} > 0 \quad (9)$$

It is clear that if (9) is satisfied, then so is (8). The above is thus a sufficient condition.

D. Uniqueness

There exist sufficient conditions ensuring that the solution of (4) is unique, which involve coherences [8]. However, the condition below is much less constraining [16, p.13]:

$$R \leq M \text{ and } R(R-1) \leq \frac{K(K-1)L(L-1)}{2} \quad (10)$$

and guarantees that there exists almost surely a unique solution. Other sufficient conditions for generic uniqueness exist [17], [18], but may be less attractive when one dimension (*i.e.* M) is large.

IV. DOA ESTIMATION

A. Optimization

The constrained optimization is carried out with the help of gradient descent type algorithms, which handle constraints in different manners. Denote for conciseness $\mathbf{x} = \text{vec}\{[\mathbf{A}^T, \mathbf{B}^T, \mathbf{S}^T]\}$ and define the objective function:

$$\mathcal{F}_\rho(\mathbf{x}; \boldsymbol{\lambda}) = \Upsilon(\mathbf{x}; \boldsymbol{\Lambda}) + \eta \exp(-\gamma \mathcal{C}_\rho(\mathbf{x})) \quad (11)$$

where η is the penalty weight, γ is introduced to control the sharpness of penalty $\mathcal{C}_\rho(\mathbf{x})$, and $\boldsymbol{\lambda}$ is defined in (7) and depends on \mathbf{x} and \mathbf{t} . This leads to the algorithm below

ALGORITHM

- 1) Choose R satisfying (10).
- 2) Initialize $(\mathbf{A}(0), \mathbf{B}(0), \mathbf{S}(0))$ to matrices with unit-norm columns satisfying $\mathcal{C}_\rho > 0$.
- 3) Compute $\mathbf{G}(0)$ and $\mathbf{f}(0)$, and solve $\mathbf{G}(0) \boldsymbol{\lambda}(0) = \mathbf{f}(0)$ for $\boldsymbol{\lambda}$, according to (7)
- 4) For $k \geq 1$ and subject to a stopping criterion, do

- a) Compute the descent direction as the gradient w.r.t. \mathbf{x} :
 $\mathbf{d}(k) = -\nabla \mathcal{F}_\rho(\mathbf{x}(k-1); \boldsymbol{\lambda}(k-1))$
- b) Compute a stepsize $\ell(k)$
- c) Update $\mathbf{x}(k) = \mathbf{x}(k-1) + \ell(k) \mathbf{d}(k)$
- d) Extract the 3 blocks from $\mathbf{x}(k)$: $\mathbf{A}(k)$, $\mathbf{B}(k)$ and $\mathbf{S}(k)$
- e) Normalize the columns of $\mathbf{A}(k)$, $\mathbf{B}(k)$ and $\mathbf{S}(k)$
- f) Compute $\mathbf{G}(k)$ and $\mathbf{f}(k)$, and solve $\mathbf{G}(k) \boldsymbol{\lambda}(k) = \mathbf{f}(k)$ for $\boldsymbol{\lambda}$, according to (7).

In the algorithm, η is decreased as the reconstruction error $\Upsilon(\mathbf{x}; \boldsymbol{\Lambda})$ decreases, whereas γ is kept fixed.

We give now some gradient expressions⁴ necessary to determine the descent direction $\mathbf{d}(k)$ when \mathcal{F}_ρ is used:

$$\frac{\partial \Upsilon}{\partial \mathbf{A}} = 2 \mathbf{A} \mathbf{M}^A - 2 \mathbf{N}^A$$

with

$$M_{pq}^A \stackrel{\text{def}}{=} \sum_{jk} \lambda_p B_{jp} S_{kp} S_{kq}^* B_{jq}^* \lambda_q^*$$

$$N_{ip}^A \stackrel{\text{def}}{=} \sum_{jk} T_{ijk} B_{jp}^* S_{kp}^* \lambda_p^*$$

and

$$\frac{\partial \exp(-\gamma \mathcal{C}_\rho)}{\partial \mathbf{A}} = \frac{\gamma}{\exp(\gamma \mathcal{C}_\rho)} \mathcal{L}_\rho^A \mathbf{A} \left[(\mathbf{A}^H \mathbf{A}) \square \mathbf{Q}^A - \mathbf{I} \right]$$

where \square denotes the Hadamard entry-wise product,

$$\mathcal{L}_\rho^A \stackrel{\text{def}}{=} \left(\sum_{q < p} |\mathbf{a}_p^H \mathbf{a}_q|^{2\rho} \right)^{\frac{-1}{2\rho} - 1} \mu(\mathbf{B}, \rho)^{-1} \mu(\mathbf{S}, \rho)^{-1},$$

and $Q_{pq} \stackrel{\text{def}}{=} |\mathbf{a}_q^H \mathbf{a}_p|^{2\rho-2}$. Keep in mind that expressions above hold true because matrix \mathbf{A} has unit-norm columns. And expressions are similar for matrices \mathbf{B} and \mathbf{S} , which also have unit-norm columns.

B. Source localisation

Since in our algorithm, matrix \mathbf{A} is computed without forcing any structure, it is necessary to extract parameters of interest (namely DoAs) from the entries of \mathbf{A} . In the present framework, we consider subarrays that are formed of linearly equispaced sensors with spacing Δ . So the elements of \mathbf{A} are given by:

$$A_{kr} = \exp \left\{ j \frac{2\pi}{\lambda} (k-1) \Delta \cos \psi_r \right\} \quad (12)$$

⁴Matrix gradients are written with the conventions described in [19], [20].

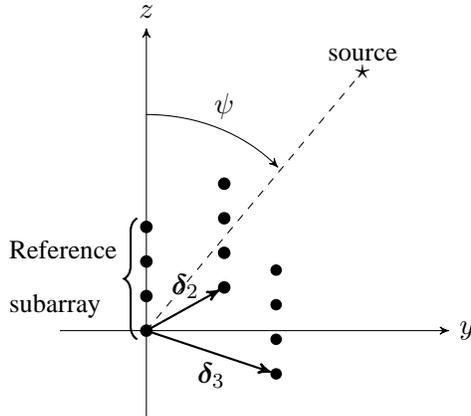


Fig. 1. One source ($R = 1$) radiating on a sensor array with $L = 3$ subarrays.

where $\lambda = \omega/2\pi C$ is the wavelength and ψ_r is the DoA of the r th source as illustrated in Figure 1. In this case, \mathbf{A} is a Vandermonde matrix with generators $\nu_r = e^{j\frac{2\pi}{\lambda}\Delta \cos \psi_r}$, $r = 1, \dots, R$. In [21], the generators of an estimated Vandermonde matrix are extracted using the following expression: $\nu_r = \frac{1}{L-1} \sum_{k=1}^{K-1} \frac{A_{k+1,r}}{A_{k,r}}$. The same idea, *i.e.*, computing the average over several estimates of the generator, is used in [22]. We propose to extract the generator set using instead a least squares minimization. Let $\underline{\mathbf{a}}_r = [A_{1,r}, \dots, A_{K-1,r}]^\top$ and $\bar{\mathbf{a}}_r = [A_{2,r}, \dots, A_{K,r}]^\top$. Then we have, $\bar{\mathbf{a}}_r \cong \nu_r \underline{\mathbf{a}}_r$. Hence the generators are optimal solutions of the following minimizations:

$$\min_v \|\bar{\mathbf{a}}_r - v \underline{\mathbf{a}}_r\|^2 \quad (13)$$

Hence $\hat{\nu}_r$ can be obtained by canceling the gradient of the previous error w.r.t v , which yields :

$$\hat{\nu}_r = \frac{\underline{\mathbf{a}}_r^H \bar{\mathbf{a}}_r}{\|\underline{\mathbf{a}}_r\|^2}. \quad (14)$$

V. COMPLEX CRAMÉR-RAO BOUNDS

When parameters are complex, expressions of Cramér-Rao bounds (CRB) depend on the definition of the complex derivative. Since a real function is never holomorphic (unless it is constant) [19], this definition is necessary; this has been overlooked in [23] but clarified in [9]. Originally, the derivative of a real function $\mathbf{h}(\boldsymbol{\theta}) \in \mathbb{R}^p$ with respect to a complex variable $\boldsymbol{\theta} \in \mathbb{C}^n$, $\boldsymbol{\theta} = \boldsymbol{\alpha} + j\boldsymbol{\beta}$, $\boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathbb{R}^n$, has been defined as the $p \times n$ matrix [19]:

$$\frac{\partial \mathbf{h}}{\partial \boldsymbol{\theta}} \stackrel{\text{def}}{=} \frac{\partial \mathbf{h}}{\partial \boldsymbol{\alpha}} + j \frac{\partial \mathbf{h}}{\partial \boldsymbol{\beta}}$$

Even if the numerical results are independent of the definition assumed for theoretical calculations, we shall subsequently assume the definition proposed in [20], for consistency with [9]:

$$\frac{\partial \mathbf{h}}{\partial \boldsymbol{\theta}} \stackrel{\text{def}}{=} \frac{1}{2} \frac{\partial \mathbf{h}}{\partial \boldsymbol{\alpha}} - \frac{j}{2} \frac{\partial \mathbf{h}}{\partial \boldsymbol{\beta}} \quad (15)$$

With this definition, one has for instance that $\partial \boldsymbol{\alpha} / \partial \boldsymbol{\theta} = \frac{1}{2} \mathbf{I}$, and $\partial \boldsymbol{\beta} / \partial \boldsymbol{\theta} = -\frac{j}{2} \mathbf{I}$. This is a key difference with [19], where we had instead: $\partial \boldsymbol{\alpha} / \partial \boldsymbol{\theta} = \mathbf{I}$, and $\partial \boldsymbol{\beta} / \partial \boldsymbol{\theta} = j \mathbf{I}$. Assume that parameter $\boldsymbol{\theta}$ is wished to be estimated from an observation \mathbf{z} , of probability distribution $p(\mathbf{z}; \boldsymbol{\theta})$, and denote $\mathbf{u}(\mathbf{z}; \boldsymbol{\theta})$ the score function. Then we have for any function $\mathbf{h}(\boldsymbol{\theta}) \in \mathbb{R}^p$:

$$\mathbb{E}\{\mathbf{h}(\mathbf{z})\mathbf{u}(\mathbf{z}; \boldsymbol{\theta})^\top\} = \frac{\partial}{\partial \boldsymbol{\theta}} \mathbb{E}\{\mathbf{h}(z)\}, \quad \text{with } \mathbf{u}(\mathbf{z}; \boldsymbol{\theta})^\top \stackrel{\text{def}}{=} \frac{\partial}{\partial \boldsymbol{\theta}} \log \mathcal{L}(\mathbf{z}; \boldsymbol{\theta}) \quad (16)$$

This is a direct consequence of the fact that $\mathbb{E}\{\mathbf{u}\} = \mathbf{0}$, valid if derivation with respect to $\boldsymbol{\theta}$ and integration with respect to $\Re(\mathbf{z})$ and $\Im(\mathbf{z})$ can be permuted. Now let $\mathbf{t}(\mathbf{z})$ be an unbiased estimator of $\boldsymbol{\theta}$. Then, following [19], one can prove that $\mathbb{E}\{\mathbf{t}\mathbf{u}^\top\} = \mathbb{E}\{(\mathbf{t} - \boldsymbol{\theta})\mathbf{u}^\top\} = \mathbf{I}$ and $\mathbb{E}\{\mathbf{t}\mathbf{u}^H\} = \mathbf{0}$. Finally, by expanding the covariance matrix of the random vector $(\mathbf{t} - \boldsymbol{\theta}) - \mathbf{F}^{-1}\mathbf{u}^*$, one readily obtains that:

$$\mathbf{V} \geq \mathbf{F}^{-1}, \quad \text{with } \mathbf{V} \stackrel{\text{def}}{=} \mathbb{E}\{(\mathbf{t} - \boldsymbol{\theta})(\mathbf{t} - \boldsymbol{\theta})^H\} \text{ and } \mathbf{F} \stackrel{\text{def}}{=} \mathbb{E}\{\mathbf{u}^* \mathbf{u}^\top\} \quad (17)$$

Note that the definition of the Fisher information matrix is the complex conjugate of that of [19], because of a different definition of the complex derivation (and hence a different definition of the complex score function). Only notations differ, and bounds on variances remain eventually the same.

VI. CRAMÉR-RAO BOUNDS OF THE LOCALIZATION PROBLEM

A. Likelihood

The noise $\mathbf{n} = \mathbf{n}_x + j\mathbf{n}_y$ is assumed to follow a complex normal distribution with zero mean. The likelihood function of \mathbf{t} , defined in (4), takes the form:

$$p(\mathbf{t}, \mathbf{t}^*) \stackrel{\text{def}}{=} \pi^{-KLM} (\det(\boldsymbol{\Sigma}) \det(\mathbf{P}))^{-1/2} \cdot \exp \left\{ -(\mathbf{t} - \boldsymbol{\mu})^H (\mathbf{P}^{-1})^* (\mathbf{t} - \boldsymbol{\mu}) + \text{Re} \left((\mathbf{t} - \boldsymbol{\mu})^\top \mathbf{R}^\top (\mathbf{P}^{-1})^* (\mathbf{t} - \boldsymbol{\mu}) \right) \right\} \quad (18)$$

where $\boldsymbol{\mu}$ is the noise free part of \mathbf{t} , $\boldsymbol{\Sigma}$ is the covariance matrix of \mathbf{t} , $\mathbf{C} = \mathbb{E}\{(\mathbf{t} - \boldsymbol{\mu})(\mathbf{t} - \boldsymbol{\mu})^\top\}$ is the noncircular covariance (sometimes called *relation matrix*), and

$$\mathbf{P} = \boldsymbol{\Sigma}^* - \mathbf{C}^H \boldsymbol{\Sigma}^{-1} \mathbf{C}, \quad \mathbf{R} = \mathbf{C}^H \boldsymbol{\Sigma}^{-1} \quad (19)$$

In our computer experiments, a general noise form will be considered, in which non-circularity will be controlled using a variable ε :

$$\text{cov} \left\{ [\mathbf{n}_x^\top, \mathbf{n}_y^\top]^\top \right\} = \begin{bmatrix} \boldsymbol{\Sigma}_{xx} & \boldsymbol{\Sigma}_{xy} \\ \boldsymbol{\Sigma}_{yx} & \boldsymbol{\Sigma}_{yy} \end{bmatrix} = \frac{\sigma^2}{2} \begin{bmatrix} (1 + \varepsilon)\mathbf{I} & \varepsilon\mathbf{I} \\ \varepsilon\mathbf{I} & (1 - \varepsilon)\mathbf{I} \end{bmatrix} \quad (20)$$

In this case $\boldsymbol{\Sigma} = \sigma^2\mathbf{I}$ and $\mathbf{C} = \sigma^2\varepsilon(1 + j)\mathbf{I}$. Therefore, the likelihood function becomes

$$p(\mathbf{t}, \mathbf{t}^*) = \left(\sigma^2 \pi (1 - 2\varepsilon^2)^{1/2} \right)^{-KLM} \cdot \exp \left\{ \frac{-1}{\sigma^2(1 - 2\varepsilon^2)} (\mathbf{t} - \boldsymbol{\mu})^H (\mathbf{t} - \boldsymbol{\mu}) + \text{Re} \left(\frac{\varepsilon(1 - j)}{\sigma^2(1 - 2\varepsilon^2)} (\mathbf{t} - \boldsymbol{\mu})^\top (\mathbf{t} - \boldsymbol{\mu}) \right) \right\} \quad (21)$$

Now, let

$$\boldsymbol{\theta} = \underbrace{[\psi_1, \dots, \psi_R]}_{\boldsymbol{\psi}}, \underbrace{[\bar{\mathbf{b}}_1^\top, \dots, \bar{\mathbf{b}}_R^\top, \mathbf{s}_1^\top, \dots, \mathbf{s}_R^\top]}_{\boldsymbol{\xi}}, \underbrace{[\bar{\mathbf{b}}_1^H, \dots, \bar{\mathbf{b}}_R^H]}_{\boldsymbol{\xi}^*} \quad (22)$$

denote the unknown parameter vector, where $\bar{\mathbf{b}}_r \stackrel{\text{def}}{=} [B_{2,r}, \dots, B_{L,r}]^\top$. Note that, by definition, the likelihood p is a function of both \mathbf{t} and \mathbf{t}^* , and not of $\|\mathbf{t} - \boldsymbol{\mu}\|$ only, especially if noise is non-circularly distributed. In this eventuality, it is hence necessary to introduce both complex parameters and their conjugates in the unknown parameter vector $\boldsymbol{\theta}$, as displayed in (22).

B. Fisher Information

Our goal now is to derive the CRBs of the parameters in $\boldsymbol{\theta}$. The CRBs for factor matrices have been computed in [9]. However, it should be emphasized that, unlike [9], no assumption is needed on the elements of matrix \mathbf{S} to derive the CRB. In fact, assuming that the first row of \mathbf{A} and \mathbf{B} is fixed to $[1, \dots, 1]_{1 \times R}$ is sufficient. Yet, the latter assumption is satisfied in the considered array configuration.

The CRB for unbiased estimation of the complex parameters $\boldsymbol{\theta}$ is equal to the inverse of the Fisher information matrix \mathbf{F} , defined in equation (17). We start with the log-likelihood:

$$\begin{aligned} \mathcal{L}(\boldsymbol{\theta}) = & -KLM \log \left(\sigma^2 \pi (1 - 2\varepsilon^2)^{1/2} \right) - \frac{1}{\sigma^2(1 - 2\varepsilon^2)} (\mathbf{t} - \boldsymbol{\mu})^H (\mathbf{t} - \boldsymbol{\mu}) \\ & + \text{Re} \left(\frac{\varepsilon(1 - j)}{\sigma^2(1 - 2\varepsilon^2)} (\mathbf{t} - \boldsymbol{\mu})^\top (\mathbf{t} - \boldsymbol{\mu}) \right) \end{aligned} \quad (23)$$

A straightforward calculation then yields:

$$\mathbf{s}^\top = \frac{1}{\sigma^2(1 - 2\varepsilon^2)} \left[\mathbf{n}^\top \frac{\partial \boldsymbol{\mu}^*}{\partial \boldsymbol{\theta}} + \mathbf{n}^H \frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\theta}} - \varepsilon \left((1 + j) \mathbf{n}^H \frac{\partial \boldsymbol{\mu}^*}{\partial \boldsymbol{\theta}} + (1 - j) \mathbf{n}^\top \frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\theta}} \right) \right] \quad (24)$$

where $\mathbf{n} = \mathbf{z} - \boldsymbol{\mu}$. By substituting the score function \mathbf{s} by its expression, the Fisher information matrix can be written as:

$$\mathbf{F} = \frac{1}{\sigma^2(1-2\varepsilon^2)} \left[\left(\frac{\partial \boldsymbol{\mu}^*}{\partial \boldsymbol{\theta}} \right)^H \left(\frac{\partial \boldsymbol{\mu}^*}{\partial \boldsymbol{\theta}} - \varepsilon(1-j) \frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\theta}} \right) + \left(\frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\theta}} \right)^H \left(\frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\theta}} - \varepsilon(1+j) \frac{\partial \boldsymbol{\mu}^*}{\partial \boldsymbol{\theta}} \right) \right] \quad (25)$$

Since parameters in $\boldsymbol{\psi}$ are real and those in $\boldsymbol{\xi}$ are complex, a first writing of the derivatives in (25) is:

$$\frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\theta}} = \left[\frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\psi}}, \quad \frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\xi}}, \quad \mathbf{0} \right] \quad \text{and} \quad \frac{\partial \boldsymbol{\mu}^*}{\partial \boldsymbol{\theta}} = \left[\left(\frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\psi}} \right)^*, \quad \mathbf{0}, \quad \left(\frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\xi}} \right)^* \right] \quad (26)$$

Therefore, the Fisher information matrix becomes:

$$\mathbf{F} = \frac{1}{\sigma^2(1-2\varepsilon^2)} \begin{bmatrix} 2 \operatorname{Re} \{ \mathbf{K}_{11} \} & \mathbf{K}_{12} & \mathbf{K}_{12}^* \\ \mathbf{K}_{12}^H & \mathbf{G}_{22} & \mathbf{H}_{22}^* \\ \mathbf{K}_{12}^T & \mathbf{H}_{22} & \mathbf{G}_{22}^* \end{bmatrix} \quad (27)$$

where

$$\mathbf{K}_{ij} = \mathbf{G}_{ij} + \mathbf{H}_{ij} \quad (28)$$

$$\mathbf{G}_{ij} = \left(\frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\theta}_i} \right)^H \left(\frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\theta}_j} \right) \quad (29)$$

$$\mathbf{H}_{ij} = \varepsilon(j-1) \left(\frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\theta}_i} \right)^T \left(\frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\theta}_j} \right) \quad (30)$$

with $(i, j) \in \{1, 2\} \times \{1, 2\}$, $\boldsymbol{\theta}_1 = \boldsymbol{\psi}$ and $\boldsymbol{\theta}_2 = \boldsymbol{\xi}$.

To complete the calculation of \mathbf{F} , it remains to give partial derivative expressions of $\boldsymbol{\mu}$ with respect to $\boldsymbol{\psi}$ and $\boldsymbol{\xi}$. This is addressed in the following subsection. To conclude, we present the Fisher information \mathbf{F}_c for a circular complex Gaussian (CCG) noise. In this case, we just need to set $\varepsilon = 0$ in (27), which yields:

$$\mathbf{F}_c = \frac{1}{\sigma^2} \begin{bmatrix} 2 \operatorname{Re} \{ \mathbf{G}_{11} \} & \mathbf{G}_{12} & \mathbf{G}_{12}^* \\ \mathbf{G}_{12}^H & \mathbf{G}_{22} & \mathbf{0} \\ \mathbf{G}_{12}^T & \mathbf{0} & \mathbf{G}_{22}^* \end{bmatrix} \quad (31)$$

In view of (31), it is clear that the introduction of $\boldsymbol{\xi}^*$ in the parameter vector is not necessary in the case where the noise follows a CCG distribution. With a non circular complex Gaussian (NCCG) noise, this is not the case.

C. Derivatives of $\boldsymbol{\mu}$ with respect to $\boldsymbol{\psi}$

Using the chain rule we have

$$\frac{\partial \boldsymbol{\mu}}{\partial \psi_f} = \left(\frac{\partial \boldsymbol{\mu}}{\partial \mathbf{a}_f^\top} \right) \left(\frac{\partial \mathbf{a}_f^\top}{\partial \psi_f} \right) \quad (32)$$

and $[\partial \boldsymbol{\mu} / \partial \mathbf{a}_f^\top]$ can be computed using complex derivative formulas. Then, we obtain:

$$\frac{\partial \boldsymbol{\mu}}{\partial \mathbf{a}_f^\top} = \mathbf{I}_K \boxtimes \mathbf{b}_f \boxtimes \mathbf{s}_f \in \mathbb{C}^{KLM \times K}, \quad 1 \leq f \leq R. \quad (33)$$

To calculate $[\partial \mathbf{a}_f^\top / \partial \psi_f]$, we use the expressions of the considered sensor array configuration, namely equation (12), which yields:

$$\frac{\partial \mathbf{a}_f^\top}{\partial \psi_f} = -j\pi \sin \psi_f (\mathbf{a}_f \square \mathbf{v}_K) \quad (34)$$

where $\mathbf{v}_K = [0, 1, \dots, K-1]^\top$. By substituting (33) and (34) in (32), we get

$$\frac{\partial \boldsymbol{\mu}}{\partial \psi_f} = -j\pi \sin \psi_f (\mathbf{I}_K \boxtimes \mathbf{b}_f \boxtimes \mathbf{s}_f) (\mathbf{a}_f \square \mathbf{v}_K) \stackrel{\text{def}}{=} \boldsymbol{\phi}_{\psi_f} \quad (35)$$

$$\text{and } \frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\psi}} = [\boldsymbol{\phi}_{\psi_1}, \dots, \boldsymbol{\phi}_{\psi_R}] \in \mathbb{C}^{KLM \times R} \quad (36)$$

D. Derivatives of $\boldsymbol{\mu}$ with respect to $\boldsymbol{\xi}$

Taking partial derivatives of $\boldsymbol{\mu}$ with respect to $\bar{\mathbf{b}}_f^\top$ and \mathbf{s}_f^\top , we obtain:

$$\frac{\partial \boldsymbol{\mu}(\boldsymbol{\theta})}{\partial \bar{\mathbf{b}}_f^\top} = (\mathbf{a}_f \boxtimes \mathbf{I}_{LM})(\mathbf{I}_L \boxtimes \mathbf{s}_f) \mathbf{J}_L \stackrel{\text{def}}{=} \boldsymbol{\phi}_{\bar{\mathbf{b}}_f} \in \mathbb{C}^{KLM \times (L-1)} \quad (37)$$

$$\frac{\partial \boldsymbol{\mu}(\boldsymbol{\theta})}{\partial \mathbf{s}_f^\top} = \mathbf{a}_f \boxtimes \mathbf{b}_f \boxtimes \mathbf{I}_M \stackrel{\text{def}}{=} \boldsymbol{\phi}_{\mathbf{s}_f} \in \mathbb{C}^{KLM \times M} \quad (38)$$

where $\mathbf{J}_L = [\mathbf{0}_{(L-1),1} \ \mathbf{I}_{L-1}]^\top \in \mathbb{C}^{L \times (L-1)}$ is a selection matrix. To sum up,

$$\frac{\partial \boldsymbol{\mu}}{\partial \boldsymbol{\xi}} = [\boldsymbol{\phi}_{\bar{\mathbf{b}}_1}, \dots, \boldsymbol{\phi}_{\bar{\mathbf{b}}_R}, \boldsymbol{\phi}_{\mathbf{s}_1}, \dots, \boldsymbol{\phi}_{\mathbf{s}_R}] \in \mathbb{C}^{KLM \times R(L+M-1)} \quad (39)$$

E. DoA Cramér-Rao bound

The CRB related to DoAs only is obtained as the first leading $R \times R$ block in matrix \mathbf{F}^{-1} , where \mathbf{F} is defined in (27). Doing this assumes that translations $\boldsymbol{\delta}_\ell$ are nuisance parameters, *i.e.* unknown but not of interest. This assumption is relevant in various applications. For instance, consider sonar buoys left floating on the surface and equipped with a device permitting to maintain their orientation towards North. The shape and orientation of subarrays mounted on each buoy are known, but relative locations of buoys are unknown. Similar examples can be encountered when installing arrays of sensors far apart, *e.g.*

when performing records on glaciers in difficult conditions: only locations of sensors within subarrays are known accurately. This realistic context has not been considered in the literature. Note that the CRB of the DoA where locations of all sensors are known can be found in [24], [25].

VII. COMPUTER RESULTS

A. Advantage of the constraint $\mathcal{C}_\rho(\mathbf{x})$

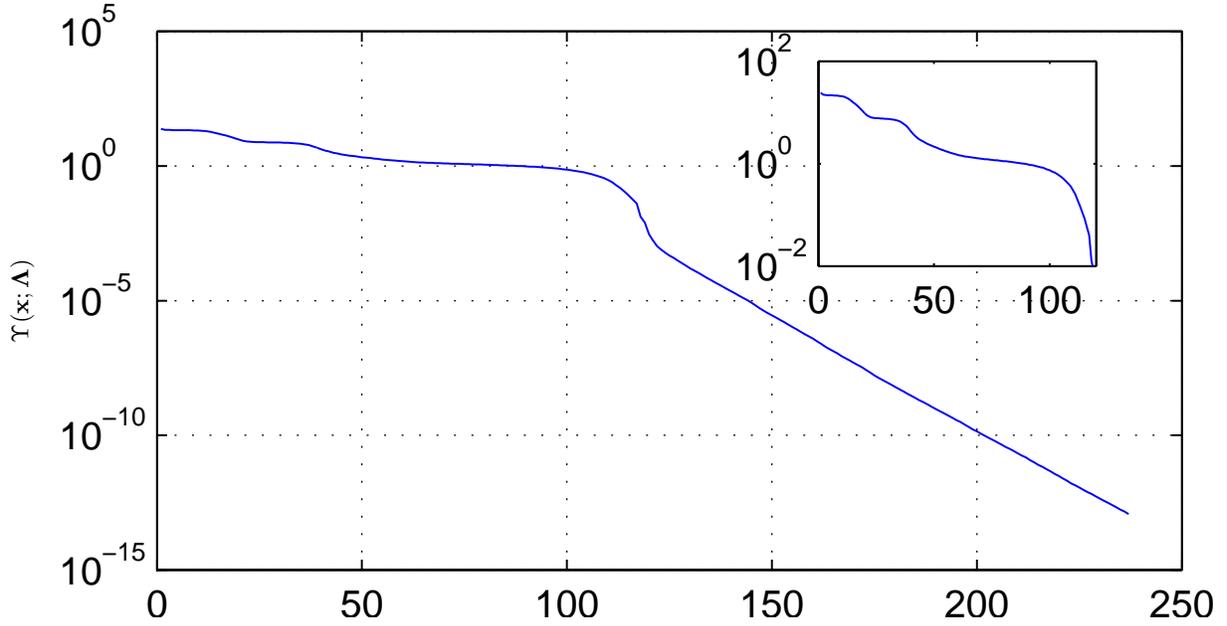
To see the interest of constraint $\mathcal{C}_\rho(\mathbf{x})$ used in the optimization algorithm, Figure 2 sketches the evolution of the reconstruction error $\Upsilon(\mathbf{x}; \mathbf{\Lambda})$, and $\mathcal{C}_\rho(\mathbf{x})$ as a function of iterations. The figure shows that thanks to the constraint \mathcal{C}_ρ : (i) iterates are incited to remain inside (or turn back into) the feasible region (where existence is guaranteed), (ii) the optimization algorithm converges quickly because iterates are allowed to move away from the feasible region (depending on parameters η and γ).

B. Monte Carlo experiments

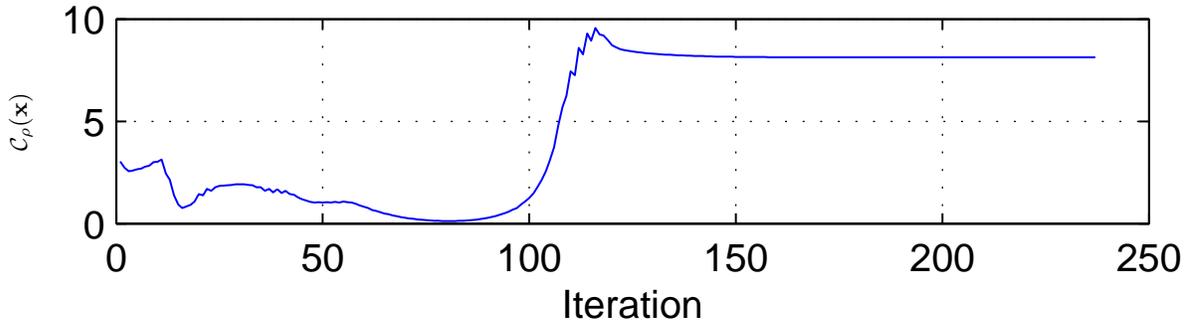
To evaluate the efficiency of the proposed method, we compare its performances to two other algorithms, ESPRIT and MUSIC [24], [25]. The performance criterion is the *total* mean square error (total MSE) of the DoA: $\frac{1}{RN} \sum_{r=1}^R \sum_{n=1}^N (\hat{\psi}_{r,n} - \psi_r)^2$ where $\hat{\psi}_{r,n}$ is the estimated DoA at the n -th Monte-Carlo trial and N is the number of trials. The deterministic CRB computed in the previous section is reported as a benchmark. The scenario on which the proposed algorithm is tested can be of interest in numerous applications, where translations δ_ℓ are unknown, as pointed out in Section VI-E. To show the influence of various parameters of the problem on the estimation results, we study four examples, whose parameters are reported in the table below:

	Subarrays	Noise	DoA
Example 1	$L = 2, (\delta_2)$	CCG	$40^\circ, 64^\circ, 83^\circ$
Example 2	$L = 3, (\delta_2, \delta_3)$	CCG	$40^\circ, 64^\circ, 83^\circ$
Example 3	$L = 3, (\delta_2, \delta_3)$	CCG	$7^\circ, 64^\circ, 83^\circ$
Example 4	$L = 3, (\delta_2, \delta_3)$	NCCG	$7^\circ, 60^\circ, 70^\circ$

where $\delta_2 = [0, 25\lambda, 0]^\top$, $\delta_3 = [0, 37.5\lambda, 5\lambda]^\top$. In all examples, each subarray is an uniform linear array (ULA) of 4-element with half-wavelength spacing (see Figure 1), and the narrowband source signals have the same power. In all experiments, $M = 200$ time samples are used, and 200 Monte-Carlo simulations are run for each SNR level. Figures 3, 4, 5 and 6 report the MSE of the DoAs obtained in Examples 1, 2, 3 and 4, respectively.



(a) Error $\Upsilon(\mathbf{x}; \Lambda)$. The insert shows a zoom between iterations 0 and 120



(b) Constraint $C_{\rho}(\mathbf{x})$

Fig. 2. Reconstruction error $\Upsilon(\mathbf{x}; \Lambda)$ and $C_{\rho}(\mathbf{x})$ as a function of the number of iterations. This is a typical example among the Monte-Carlo experiments that have been run.

Example 1: This experiment shows that: (i) the proposed CP algorithm exhibits the same performances as ESPRIT, which makes sense, (ii) MUSIC performs the best, but exploits more information, namely the exact knowledge of sensor locations, whereas this information is actually not available in the present scenario. Hence MUSIC performances just serve as a reference.

Example 2: This experiment shows that the proposed algorithm yields better results than EPSRIT. The reason is that ESPRIT uses at most two subarrays, whereas the proposed algorithm uses all of them. Again, MUSIC is reported just as a reference benchmark.

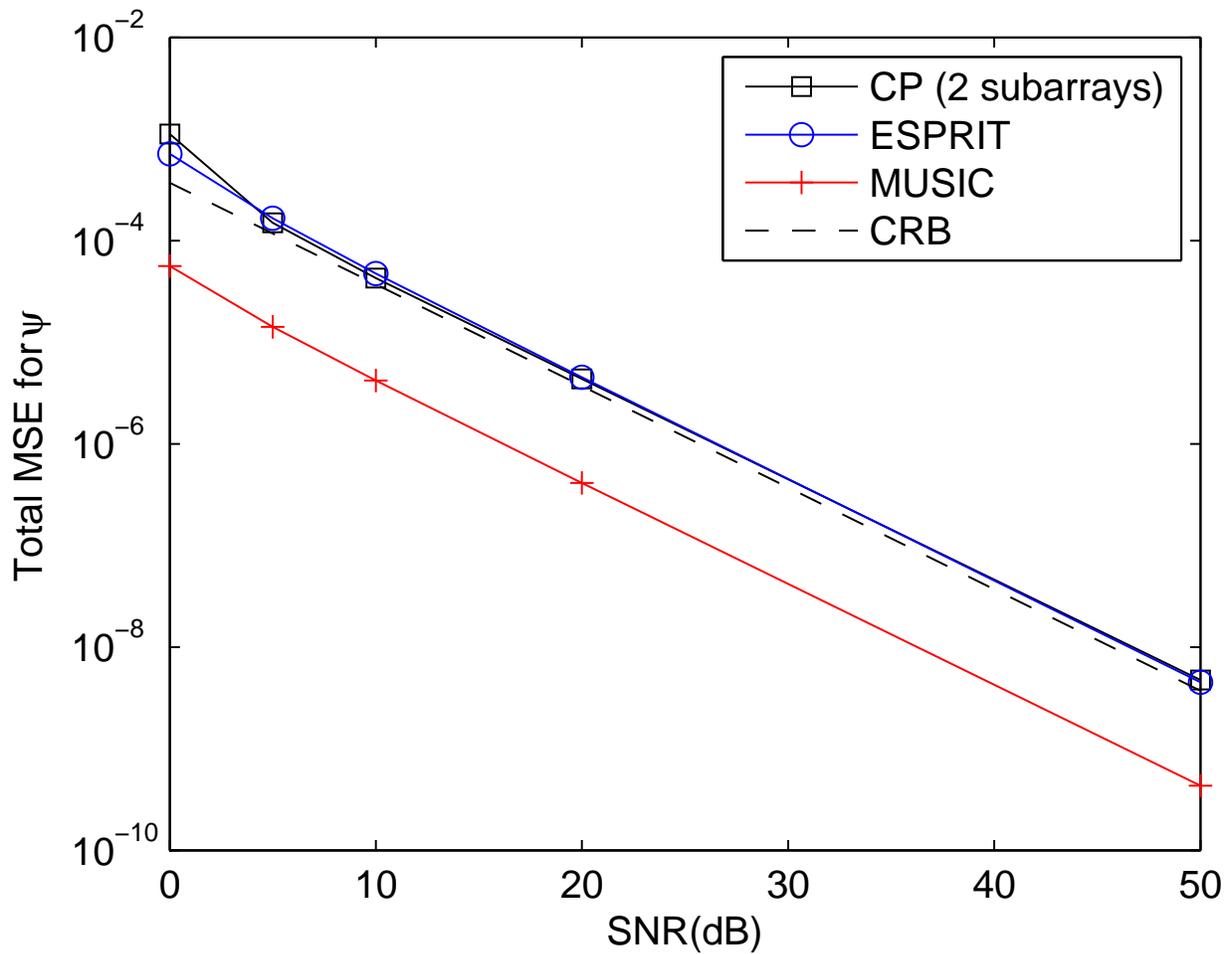


Fig. 3. Total DoA error versus SNR, with $L = 2$ subarrays, $\psi = [40^\circ, 65^\circ, 83^\circ]$.

Example 3: This experiment shows the same results as in example 2, except for an increase in MSE at low SNR, which is due to the direction of arrival $\psi = 7^\circ$. Actually, for an ULA, the source localization accuracy degrades as the DoAs come closer to the end-fire, so that the so-called *threshold region* [26], [27] (which always exists at low SNR) becomes visible.

Example 4: Unlike previous experiments, Monte Carlo runs in this example are performed with non-circular complex Gaussian noise. The obtained results are similar to those obtained in Examples 2 and 3.

To show the influence of the noise distribution on DoA estimation, we plot the results obtained by varying ε using the same array configuration and DoA parameters as in Example 4. Figure 7 presents CRB for different values of ε . This shows that the CRB changes with ε . Figure 8 depicts results comparing

total MSE obtained by the proposed algorithm to the CRB, each subfigure corresponding to one value of ε . We conclude that the proposed algorithm yields accurate estimates in the presence of circular as well as noncircular additive Gaussian noise.

VIII. CONCLUSION

DoA estimation of narrow-band far-field sources is formulated as a CP decomposition, when sensor arrays consist of L identical displaced subarrays. We proposed an optimization algorithm including a new differentiable penalty ensuring existence of the low-rank tensor approximation. We also derived the expressions of the Cramér-Rao bounds of DoA parameters in the presence of nuisance parameters estimated by CP decomposition methods. It was shown that, thanks to our penalty, the proposed algorithm converges quickly and is prevented to leave for long the feasible region. As expected, DoA estimation results show that the CP algorithm exhibits better results than reference DoA estimation methods when $L > 2$.

REFERENCES

- [1] H. Krim and M. Viberg, "Two decades of array signal processing research," *IEEE Sig. Proc. Mag.*, pp. 67–95, July 1996.
- [2] P. Stoica, P. Babu, and J. Li, "SPICE: A sparse covariance-based estimation method for array processing," *IEEE Trans. Signal Process.*, vol. 59, no. 2, pp. 629–638, 2011.
- [3] S. Sahnoun, E.-H. Djermoune, C. Soussen, and D. Brie, "Sparse multidimensional modal analysis using a multigrid dictionary refinement," *EURASIP J. Adv. Signal Process.*, March 2012.
- [4] R. O. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antenna Propagation*, vol. 34, no. 3, pp. 276–280, Mar. 1986.
- [5] H. L. Van Trees, *Optimum Array Processing*, vol. IV, Wiley, New York, 2002.
- [6] R. Roy and T. Kailath, "ESPRIT - estimation of signal parameters via rotational invariance techniques," *IEEE Trans. Acoust. Speech Signal Proc.*, vol. 37, pp. 984–995, July 1989.
- [7] N. D. Sidiropoulos, R. Bro, and G. B. Giannakis, "Parallel factor analysis in sensor array processing," *IEEE Trans. Sig. Proc.*, vol. 48, no. 8, pp. 2377–2388, Aug. 2000.
- [8] L.-H. Lim and P. Comon, "Blind multilinear identification," *IEEE Trans. Inf. Theory*, vol. 60, no. 2, pp. 1260–1280, Feb. 2014, open access.
- [9] X. Liu and N. Sidiropoulos, "Cramér-Rao lower bounds for low-rank decomposition of multidimensional arrays," *IEEE Trans. Sig. Proc.*, vol. 49, no. 9, pp. 2074–2086, 2001.
- [10] P. Tichavsky, A. H. Phan, and Z. Koldovsky, "Cramér-rao-induced bounds for Candecomp/Parafac tensor decomposition," *IEEE Transactions on Signal Processing*, 2013.
- [11] P. Comon, "Tensors: a brief introduction," *IEEE Sig. Proc. Magazine*, vol. 31, no. 3, May 2014, special issue on BSS. hal-00923279.
- [12] H. A. L. Kiers, "Towards a standardized notation and terminology in multiway analysis," *J. Chemometrics*, pp. 105–122, 2000.

- [13] P. Comon, "Tensors, usefulness and unexpected properties," in *15th IEEE Workshop on Statistical Signal Processing (SSP'09)*, Cardiff, UK, Aug. 31 - Sep. 3 2009, pp. 781–788, keynote. hal-00417258.
- [14] R. Gribonval and M. Nielsen, "Sparse representations in unions of bases," *IEEE Trans. Information Theory*, vol. 49, no. 12, pp. 3320–3325, Dec. 2003.
- [15] D. L. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ_1 minimization," *Proc. Nat. Acad. Sci.*, vol. 100, no. 5, pp. 2197–2202, Mar. 2003.
- [16] L. De Lathauwer, "A link between canonical decomposition in multilinear algebra and simultaneous matrix diagonalization," *SIAM Journal on Matrix Analysis*, vol. 28, no. 3, pp. 642–666, 2006.
- [17] M. V. Catalisano, A. V. Geramita, and A. Gimigliano, "Higher secant varieties of the Segre varieties," *Journal of Pure and Applied Algebra*, vol. 201, no. 1-3, pp. 367–380, 2005.
- [18] H. Abo, G. Ottaviani, and C. Peterson, "Induction for secant varieties of Segre varieties," *Trans. Amer. Math. Soc.*, pp. 767–792, 2009, arXiv:math/0607191.
- [19] P. Comon, "Estimation multivariable complexe," *Traitement du Signal*, vol. 3, no. 2, pp. 97–101, Apr. 1986, hal-00979476.
- [20] A. Hjørungnes and D. Gesbert, "Complex-valued matrix differentiation: Techniques and key results," *IEEE Trans. Sig. Proc.*, vol. 55, no. 6, pp. 2740–2746, June 2007.
- [21] R. Boyer and J. Picheral, "Second-order near-field localization with automatic paring operation," in *ICASSP 2008*. IEEE, 2008, pp. 2569–2572.
- [22] J. Liu and X. Liu, "An eigenvector-based approach for multidimensional frequency estimation with improved identifiability," *IEEE Trans. Signal Process.*, vol. 54, no. 12, pp. 4543–4556, decembre 2006.
- [23] A. van den Bos, "A Cramér-Rao bound for complex parameters," *IEEE Trans. Sig. Proc.*, vol. 42, no. 10, pp. 2859, Oct. 1994.
- [24] P. Stoica and A. Nehorai, "MUSIC, maximum likelihood, and Cramer-Rao bound," *IEEE Trans. Acoust. Speech Sig. Proc.*, vol. 37, no. 5, pp. 720–741, 1989.
- [25] B. Ottersten, M. Viberg, and T. Kailath, "Performance analysis of the total least squares ESPRIT algorithm," *IEEE Trans. Sig. Proc.*, vol. 39, no. 5, pp. 1122–1135, 1991.
- [26] F. Athley, "Threshold region performance of maximum likelihood direction of arrival estimators," *IEEE Transactions on Signal Processing*, vol. 53, no. 4, pp. 1359–1373, 2005.
- [27] E. Chaumette, A. Renaux, and P. Larzabal, "New trends in deterministic lower bounds and SNR threshold estimation: from derivable bounds to conjectural bounds," in *IEEE Sensor Array and Multichannel Signal Processing Workshop (SAM)*. IEEE, 2010, pp. 121–124.

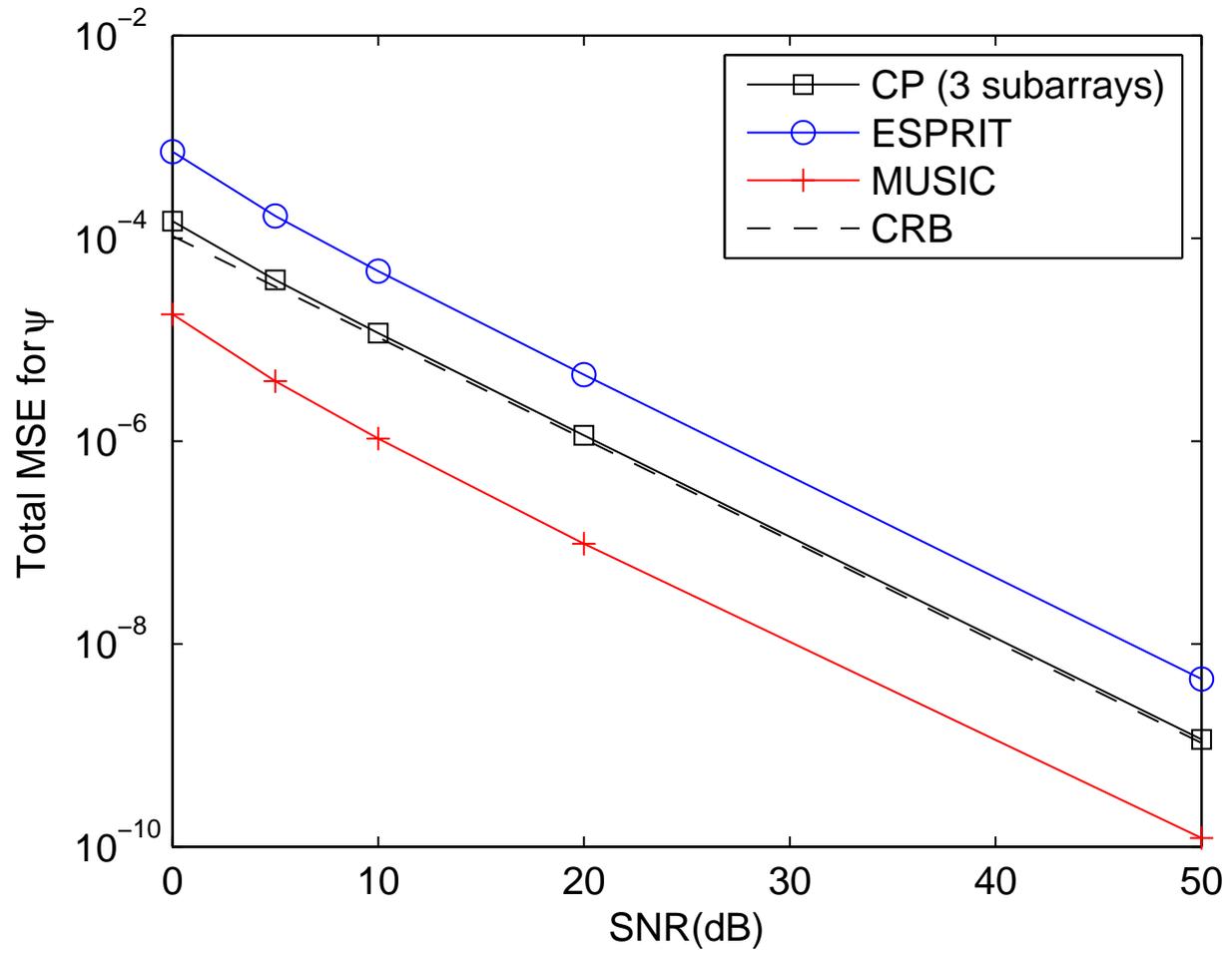


Fig. 4. Total DoA error versus SNR, with $L = 3$ subarrays, $\psi = [40^\circ, 65^\circ, 83^\circ]$.

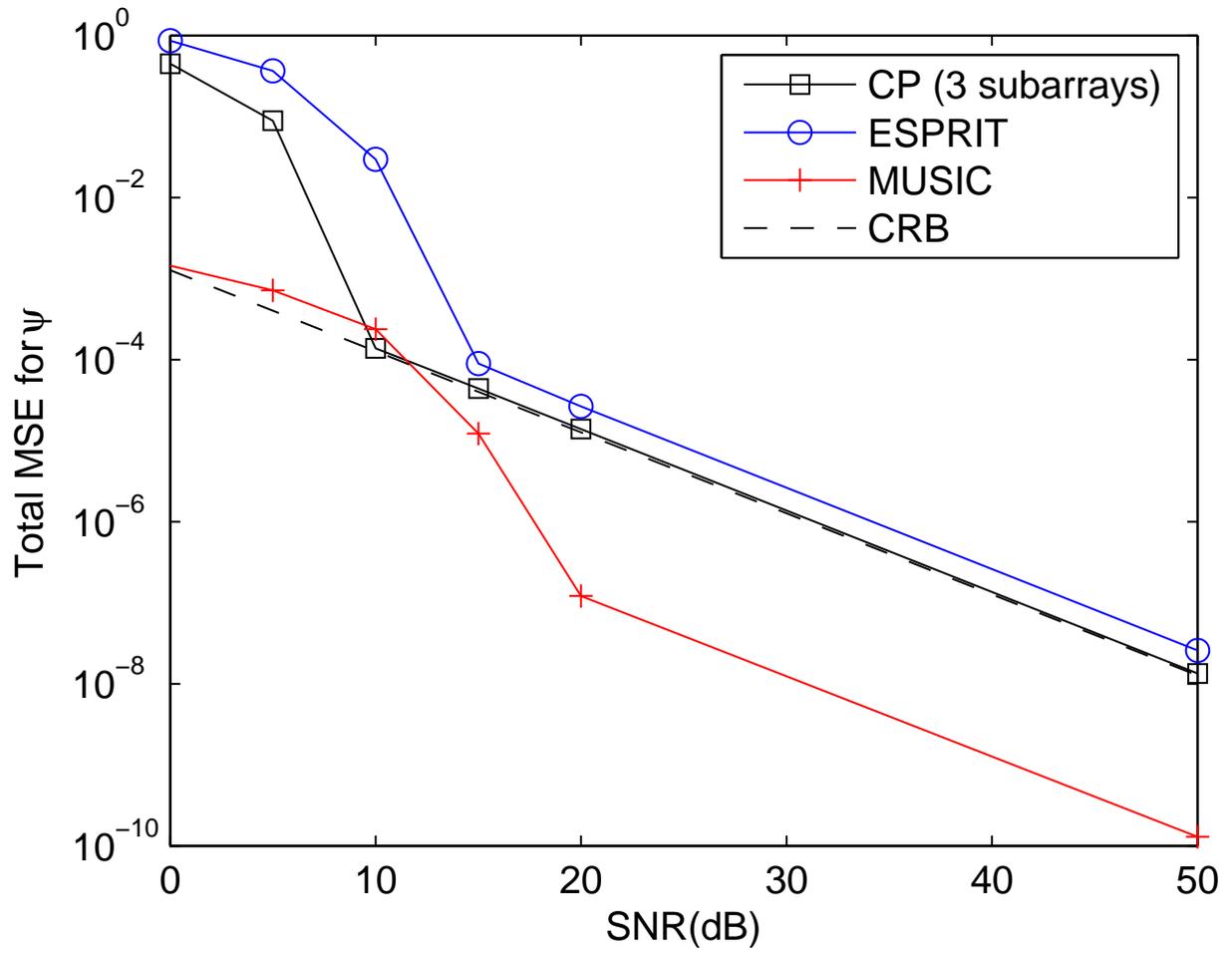


Fig. 5. Total DoA error versus SNR, with $L = 3$ subarrays, $\psi = [7^\circ, 65^\circ, 83^\circ]$.

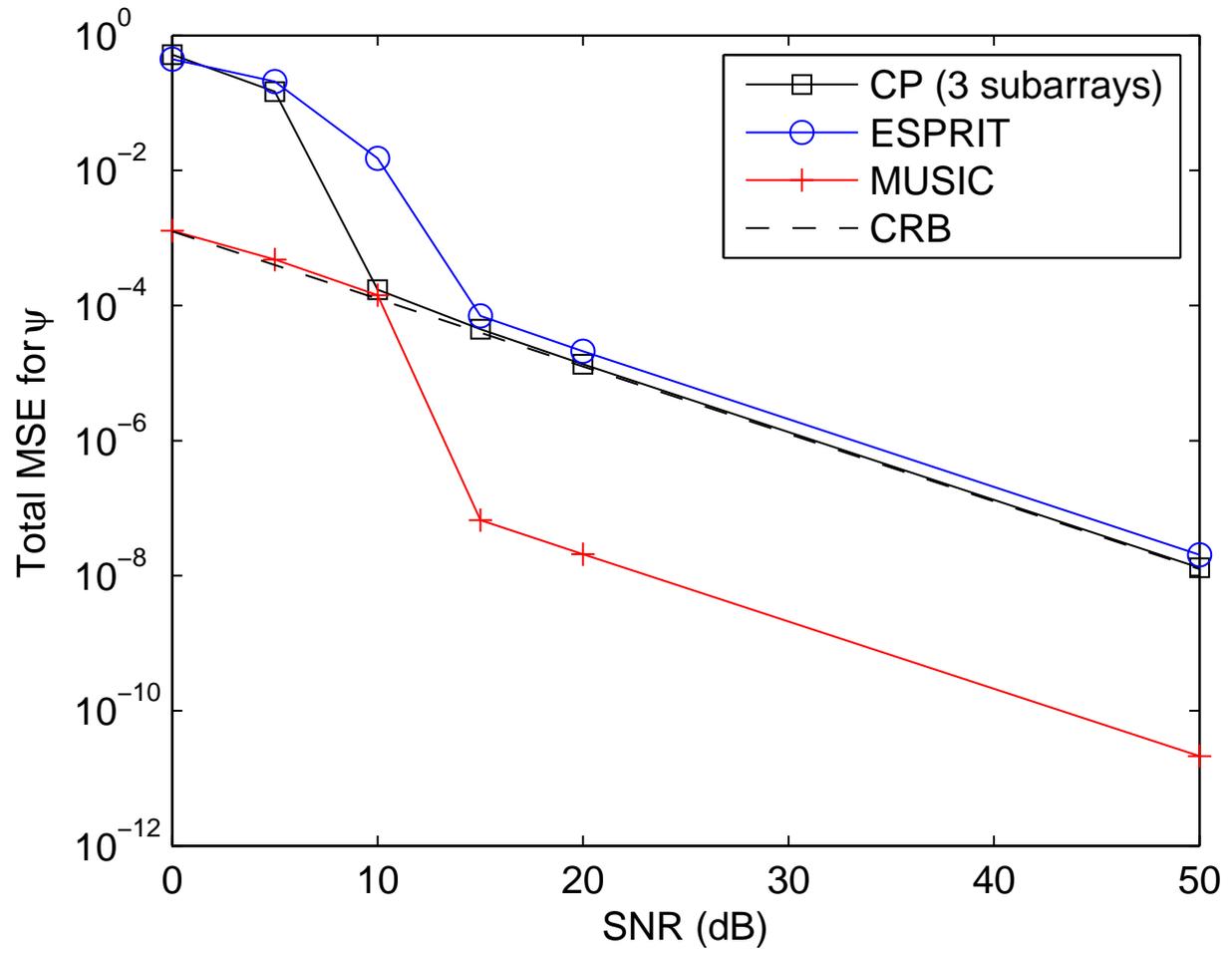


Fig. 6. Total DoA error versus SNR, with $L = 3$ subarrays, $\psi = [7^\circ, 60^\circ, 70^\circ]$, non-circular noise ($\varepsilon = 0.1$).

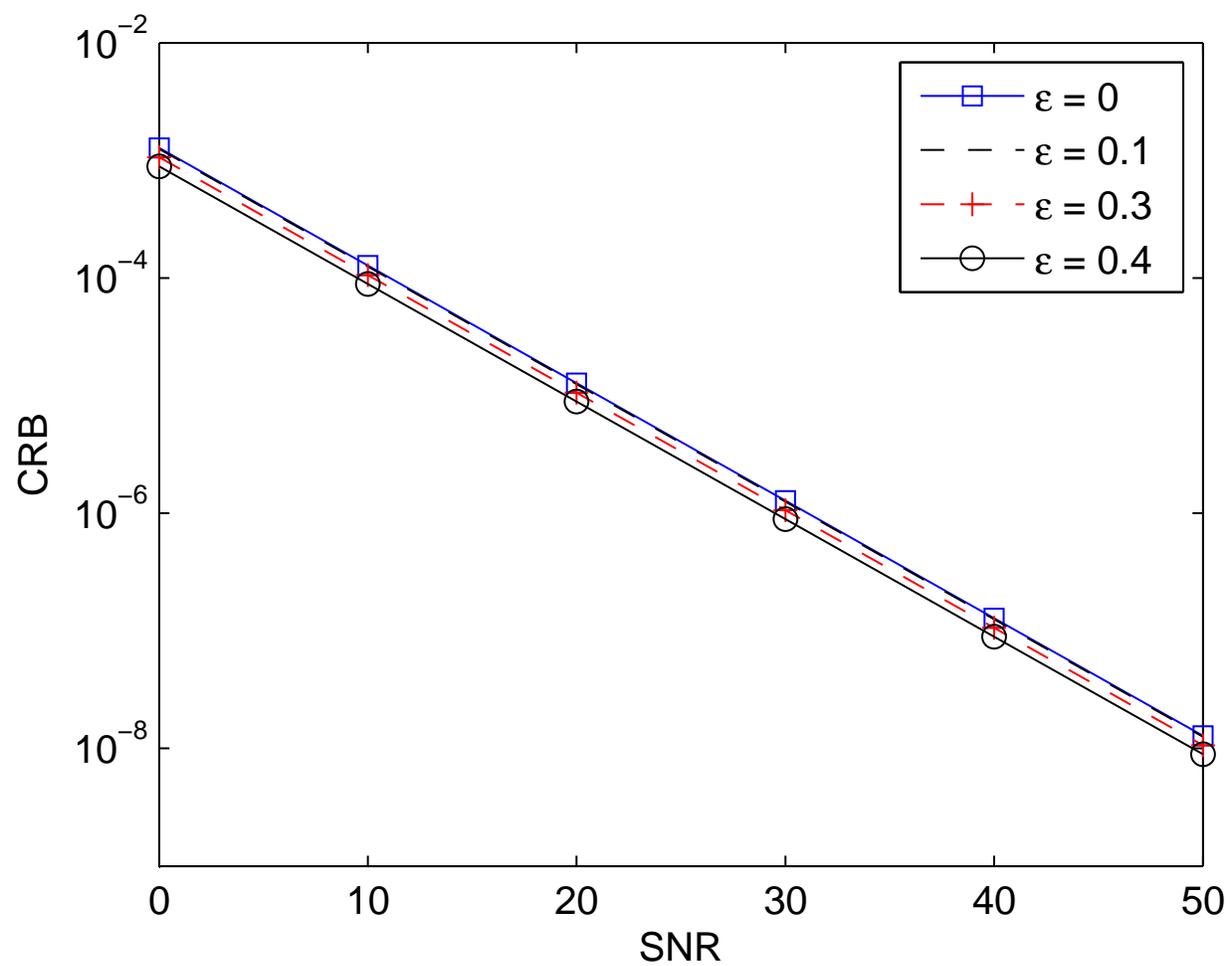


Fig. 7. CRB of ψ for different values of ε (0, 0.1, 0.3, 0.4), $L = 3$ subarrays, $\psi = [7^\circ, 60^\circ, 70^\circ]$.

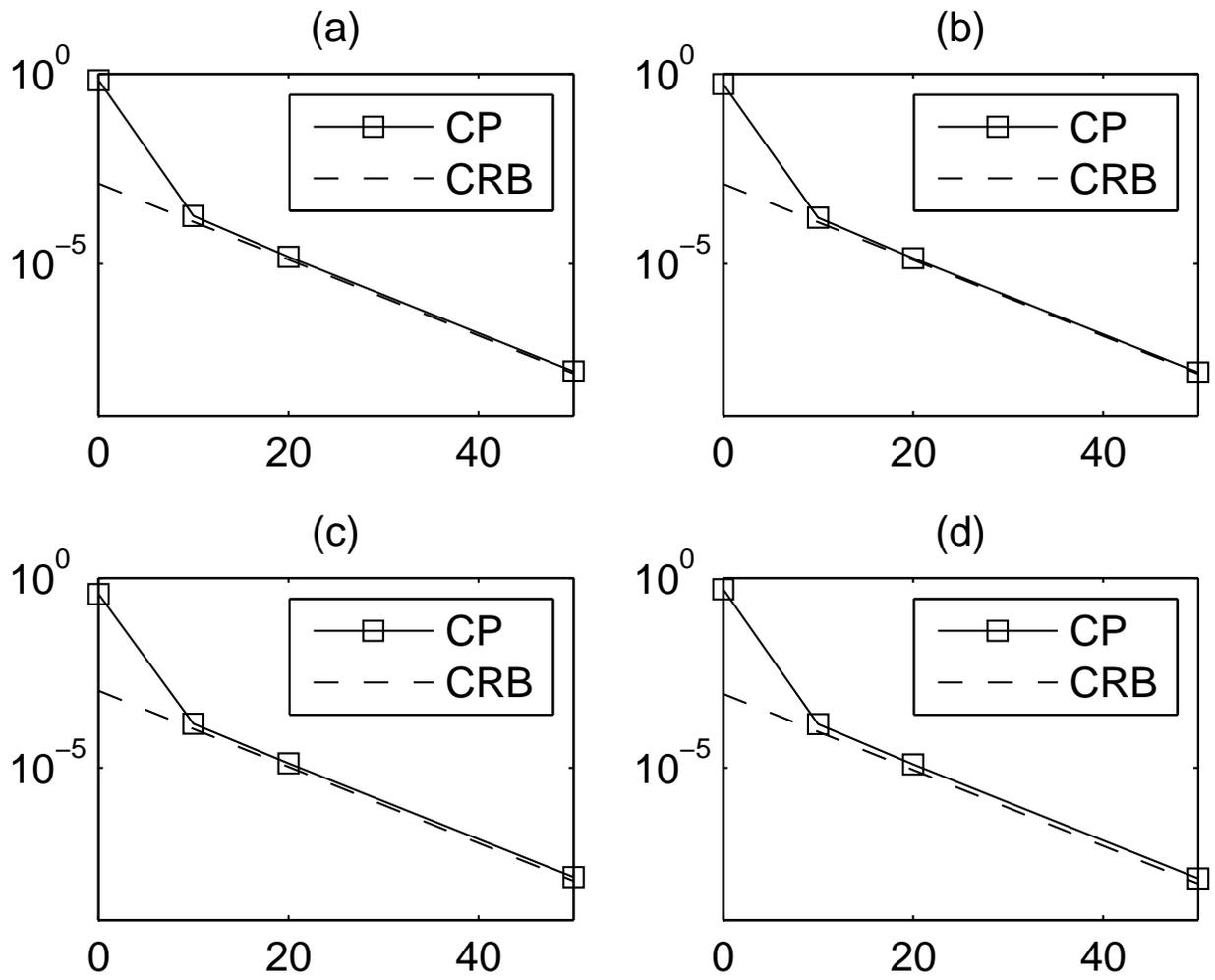


Fig. 8. Total DoA error versus SNR, with $L = 3$ subarrays, $\psi = [7^\circ, 60^\circ, 70^\circ]$. (a) $\epsilon = 0$, (b) $\epsilon = 0.1$, (c) $\epsilon = 0.3$, (d) $\epsilon = 0.4$.