Signal Processing 111 (2015) 210–221

Contents lists available at ScienceDirect

Signal Processing

journal homepage: www.elsevier.com/locate/sigpro

Survey and some new results on performance analysis of complex-valued parameter estimators

Jean Pierre Delmas [a,*], Habti Abeida [b]

[a] Institut TELECOM, TELECOM SudParis, Département CITI, CNRS UMR 5157, 91011 Evry Cedex, France
[b] Department of Electrical Engineering, University of Taif, Al-Haweiah 21974, Saudi Arabia

ARTICLE INFO

Article history:
Received 24 October 2014
Accepted 9 December 2014
Available online 26 December 2014

Keywords:
Circular (proper) and noncircular (improper) complex-valued signals
Statistical performance analysis
Cramer–Rao bound
Asymptotically minimum variance bound
Slepian–Bangs and Whittle formulas

ABSTRACT

Recently, there has been an increased awareness that simplistic adaptation of performance analysis developed for random real-valued signals and parameters to the complex case may be inadequate or may lead to intractable calculations. Unfortunately, many fundamental statistical tools for handling complex-valued parameter estimators are missing or scattered in the open literature. In this paper, we survey some known results and provide a rigorous and unified framework to study the statistical performance of complex-valued parameter estimators with a particular attention paid to properness (i.e., second order circularity), specifically referring to the second-order statistical properties. In particular, some new properties relative to the properness of the estimates, asymptotically minimum variance bound and Whittle formulas are presented. A new look at the role of nuisance parameters is given, proving and illustrating that the noncircular Gaussian distributions do not necessarily improve the Cramer–Rao bound (CRB) with respect to the circular case. Efficiency of subspace-based complex-valued parameter estimators that are presented with a special emphasis is put on noisy linear mixture.

1. Introduction

Complex-valued random signals associated with complex-valued parameters play an increasingly important role in many science and engineering problems, including those in communications, radar, biomedicine, geophysics, oceanography, electromagnetics, and optics, among others (see, e.g., [1,2] and the references therein). But the usual way to analyze the statistical performance of complex-valued parameter estimators is still often by splitting each complex parameter into its real and imaginary parts and treating them as separate real parameters [3,4]. Although this procedure is mathematically correct, it involves complicated expressions, lacking the engineering insight necessary for a lucid understanding of the various phenomena and for suggesting improved solutions. Unfortunately, many fundamental statistical tools for handling complex-valued parameter estimators are missing or scattered in open literature (see, e.g., [5, Chapter 6] and the references therein).

In this paper, we provide a rigorous and unified framework to study the statistical performance of complex-valued parameter estimators. As for all parameter estimation, an algorithm or an estimator extracts an approximation $\hat{\boldsymbol{\theta}}_N$ of an unknown parameter $\boldsymbol{\theta}$ from measurements $[\mathbf{x}(1),\ldots,\mathbf{x}(N)]$. Here the measurements are characterized by a joint PDF $p(\mathbf{x}(n)_{n=1,\ldots,N};\boldsymbol{\theta},\boldsymbol{\alpha})$ where $\mathbf{x}(n)\in\mathbb{C}^r$, $\boldsymbol{\theta}=(\theta_1,\ldots,\theta_q)^T\in\mathbb{C}^q$ is the parameter of interest and $\boldsymbol{\alpha}$ gathers all the other unknown parameters (nuisance parameters). There are two issues to consider in performance analysis. The first one, which is

* Corresponding author. Tel.: +33 1 60 76 46 32; fax: +33 1 60 76 44 33.
E-mail addresses: jean-pierre.delmas@it-sudparis.eu (J.P. Delmas), abeida3@yahoo.fr (H. Abeida).

http://dx.doi.org/10.1016/j.sigpro.2014.12.009
0165-1684/

treated in Section 2, consists in studying the performance of a particular algorithm, principally to derive the asymptotic[1] distribution, bias and covariance of $\hat{\boldsymbol{\theta}}_N$. In this section, particular attention is paid to properness (i.e., second-order circularity) of the asymptotic distribution of the parameter estimates where new properties are given. The second one is to establish a limit on the accuracy of any estimator belonging to a family of estimators. This is the subject of Sections 3 and 4, respectively, dedicated to the Cramer–Rao and asymptotically minimum variance bounds, where new extensions are given. A particular treatment of the noisy linear mixture is given in Section 5, where the efficiency of subspace-based complex-valued parameter estimators is studied. Illustration are given in Section 6 dedicated to blind identification of complex SIMO channels and complex independent component analysis. Finally, Section 7 presents a conclusion.

The following notations are used throughout the paper. Matrices and vectors are represented by bold upper case and bold lower case characters, respectively. Vectors are by default in column orientation, while T, H, $*$ and # stand for transpose, conjugate transpose, conjugate and Moore–Penrose inverse, respectively. $o(\epsilon)$ denotes a quantity such that $\lim_{\epsilon\to 0} o(\epsilon)/\epsilon = 0$. $\mathrm{E}(\cdot)$, $\mathrm{Tr}(\cdot)$, $\mathrm{Re}(\cdot)$ and $\mathrm{Im}(\cdot)$ are the expectation, trace, real and imaginary part operators, respectively. $\mathbf{I}$ is the identity matrix. $\mathrm{vec}(\cdot)$ is the "vectorization" operator that turns a matrix into a vector by stacking the columns of the matrix one below another which is used in conjunction with the Kronecker product $\mathbf{A}\otimes\mathbf{B}$ as the block matrix whose (i,j) block element is $a_{ij}\mathbf{B}$ and with the vec-permutation matrix $\mathbf{K}$ which transforms $\mathrm{vec}(\mathbf{C})$ to $\mathrm{vec}(\mathbf{C}^T)$ for any square matrix $\mathbf{C}$. $\mathrm{v}(\cdot)$ denotes the operator obtained from $\mathrm{vec}(\cdot)$ by eliminating all supradiagonal elements of the matrix. $\tilde{\mathbf{a}} = (\mathbf{a}^T, \mathbf{a}^H)^T$ and $\overline{\mathbf{a}} = (\mathrm{Re}(\mathbf{a})^T, \mathrm{Im}(\mathbf{a})^T)^T$ are, respectively, the augmented and the real-valued vector associated with complex-valued vector $\mathbf{a}$.

2. Performance analysis

To study the asymptotic performance of an algorithm, it is fruitful to adopt a functional analysis that consists in recognizing that the whole process of constructing the estimate $\hat{\boldsymbol{\theta}}_N$ is equivalent to defining a functional relation linking this estimate to the measurements from which it is inferred. Note that this functional analysis first introduced in [6] has been presented in [7, Section 3.1] for the real-valued DOA parameter and can be considered as an extension of this one. As generally $\hat{\boldsymbol{\theta}}_N$ are functions of some statistics $\mathbf{s}_N \in \mathbb{C}^p$ deduced from $(\mathbf{x}(n))_{n=1,\ldots,N}$, we have the following mapping:

$$(\mathbf{x}(n))_{n=1,\ldots,N} \longmapsto \mathbf{s}_N \xmapsto{\mathbf{g}} \hat{\boldsymbol{\theta}}_N. \tag{1}$$

The statistic $\mathbf{s}_N$ is assumed to converge almost surely to $\mathbf{s}(\boldsymbol{\theta}) = \mathrm{E}(\mathbf{s}_N)$ and $\boldsymbol{\theta}$ is supposed identifiable from $\mathbf{s}(\boldsymbol{\theta})$, so generally $p \geq q$. The functional dependence $\hat{\boldsymbol{\theta}}_N = \mathbf{g}(\mathbf{s}_N)$ constitutes an extension of the mapping $\mathbf{s}(\boldsymbol{\theta}) \longmapsto \boldsymbol{\theta}$ in the neighborhood of $\mathbf{s}(\boldsymbol{\theta})$. Each extension $\mathbf{g}$ specifies a particular algorithm. The statistics $\mathbf{s}_N$ may be sample moments,[2] cumulants of $\mathbf{x}(n)$ or any specific statistics adapted to the distribution of the measurements. For example, for the noisy linear mixtures (31), the orthogonal projectors associated with the sample estimates of the covariance, complementary covariance[3] or augmented covariance of $\mathbf{x}(n)$ have been used for subspace-based algorithms. In the specific case of independent identically distributed (IID) measurements $\mathbf{x}(n)$, closed-form expressions $(1/N)\mathbf{R}_s$ and $(1/N)\mathbf{C}_s$ of the covariance $\mathrm{E}[(\mathbf{s}_N-\mathbf{s})(\mathbf{s}_N-\mathbf{s})^H]$ and complementary covariance $\mathrm{E}[(\mathbf{s}_N-\mathbf{s})(\mathbf{s}_N-\mathbf{s})^T]$ matrices of $\mathbf{s}_N$ (where $\mathbf{s}=\mathbf{s}(\boldsymbol{\theta})$ for short) can be easily derived for sample moments or cumulants of $\mathbf{x}(n)$. For stationary measurements $\mathbf{x}(n)$ and associated sample statistics $\mathbf{s}_N$, central limit theorems and standard theorems of continuity allow us to derive convergences in distribution w.r.t. the number N of the measurements (see, e.g., [9] for the second-order statistics $\mathbf{s}_N$), viz.,

$$\sqrt{N}(\mathbf{s}_N-\mathbf{s}) \xrightarrow{\mathcal{L}} \mathcal{N}_C(\mathbf{0};\mathbf{R}_s,\mathbf{C}_s), \tag{2}$$

where $\mathcal{N}_C(\mathbf{m};\mathbf{R},\mathbf{C})$ denotes the complex Gaussian distribution with mean $\mathbf{m}$, covariance $\mathbf{R}$ and complementary covariance $\mathbf{C}$, defined as the distribution of a complex-valued random variable $\mathbf{z}$ such that the associated scalar real-valued random variable $\tilde{\mathbf{a}}^H\tilde{\mathbf{z}}$ is Gaussian distributed $\mathcal{N}_R(\mathbf{m}_z;\mathbf{R}_z)$ with mean $\mathbf{m}_z = \tilde{\mathbf{a}}^H\tilde{\mathbf{m}}$ and covariance $\mathbf{R}_z = \tilde{\mathbf{a}}^H\left(\begin{smallmatrix}\mathbf{R} & \mathbf{C}\\ \mathbf{C}^* & \mathbf{R}^*\end{smallmatrix}\right)\tilde{\mathbf{a}}$ for any vector $\mathbf{a}$ of compatible dimension.

Finally, note that this functional analysis (1) is not always relevant for other distributions of $(\mathbf{x}(n))_{n=1,\ldots,N}$. For example, if $\mathbf{x}(n) = \mathbf{s}(\boldsymbol{\theta},n) + \mathbf{e}(n)$ where $\mathbf{s}(\boldsymbol{\theta},n)$ is a nonlinear deterministic function[4] of $\boldsymbol{\theta}$ and $(\mathbf{e}(n))_{n=1,\ldots,N}$ is the zero-mean IID, $(\mathbf{x}(n))_{n=1,\ldots,N}$ are independent, but not identically distributed and the speed of convergence of the sequence of estimates $\hat{\boldsymbol{\theta}}_N$ can depend on its component and be different from $\sqrt{N}$.

For statistics $\mathbf{s}_N$ satisfying (2) and for $\mathbb{R}$-differentiable mapping $\mathbf{g}$ [10],

$$\mathbf{g}(\mathbf{s}_N) = \mathbf{g}(\mathbf{s}) + \mathbf{D}_g(\mathbf{s}_N-\mathbf{s}) + \mathbf{D}_{*g}(\mathbf{s}_N-\mathbf{s})^* + o(\|\mathbf{s}_N-\mathbf{s}\|), \tag{3}$$

where $\mathbf{D}_g$ and $\mathbf{D}_{*g}$, $q\times p$ matrices, denote the $\mathbb{R}$-derivative $\partial\mathbf{g}/\partial\mathbf{s}$ and the conjugate $\mathbb{R}$-derivative $\partial\mathbf{g}/\partial\mathbf{s}^*$ of $\mathbf{g}$ at point $\mathbf{s}$ [11]. In practice, the matrices $\mathbf{D}_g$ and $\mathbf{D}_{*g}$ are derived from perturbation analysis where only the wide-linear term is kept. Furthermore, their derivations are simplified from the chain rule by decomposing the mapping $\mathbf{g}$ (i.e., the algorithm) as successive simpler mappings.

It is proved in the Appendix that if $[\mathbf{D}_g, \mathbf{D}_{*g}] \neq \mathbf{0}$, the following convergence in distribution holds:

$$\sqrt{N}(\hat{\boldsymbol{\theta}}_N-\boldsymbol{\theta}) \xrightarrow{\mathcal{L}} \mathcal{N}_C(\mathbf{0};\mathbf{R}_\theta,\mathbf{C}_\theta) \tag{4}$$

[1] In general only the asymptotic distribution, bias and covariance can be derived, either w.r.t. the number N of measurements, the size r or the signal to noise ratio of the measurements. Hopefully, in practice the obtained results give good approximations for finite values of these quantities.

[2] This is typically the case for the estimates derived from the method of moments, where $\mathbf{s}_N$ are the sample moments of $\mathbf{x}(n)$. This is also the case for the maximum likelihood estimator for Gaussian distribution for which $\mathbf{s}_N$ are first and second-order statistics.

[3] Other names for complementary covariance matrix include pseudo-covariance matrix, conjugate covariance matrix and relation matrix.

[4] The celebrated noisy sinusoid case where $\mathbf{s}(\boldsymbol{\theta},n) = \sum_{k=1}^{K} a_k e^{i(2\pi n f_k+\phi_k)}$ is such an example.

with

$$\mathbf{R}_\theta = [\mathbf{D}_g, \mathbf{D}_{*g}] \begin{bmatrix} \mathbf{R}_s & \mathbf{C}_s \\ \mathbf{C}_s^* & \mathbf{R}_s^* \end{bmatrix} \begin{bmatrix} \mathbf{D}_g^H \\ \mathbf{D}_{*g}^H \end{bmatrix} \quad \text{and} \quad \mathbf{C}_\theta = [\mathbf{D}_g, \mathbf{D}_{*g}] \begin{bmatrix} \mathbf{C}_s & \mathbf{R}_s \\ \mathbf{R}_s^* & \mathbf{C}_s^* \end{bmatrix} \begin{bmatrix} \mathbf{D}_g^T \\ \mathbf{D}_{*g}^T \end{bmatrix}. \tag{5}$$

From (5), we deduce that for $\mathbf{g}$ $\mathbb{C}$-differentiable at point $\mathbf{s}$, $\mathbf{D}_{*g} = \mathbf{0}$, and the usual expressions

$$\mathbf{R}_\theta = \mathbf{D}_g \mathbf{R}_s \mathbf{D}_g^H \quad \text{and} \quad \mathbf{C}_\theta = \mathbf{D}_g \mathbf{C}_s \mathbf{D}_g^T \tag{6}$$

are derived. Furthermore in this case $\hat{\boldsymbol{\theta}}_N$ is asymptotically proper (i.e., $\mathbf{C}_\theta = \mathbf{0}$) if and only if $\mathbf{s}_N$ is asymptotically proper (i.e., $\mathbf{C}_s = \mathbf{0}$). We note that generally, if $\mathbf{s}_N$ is asymptotically proper, $\hat{\boldsymbol{\theta}}_N$ is not necessarily asymptotically proper. It becomes proper if $\mathbf{D}_g^* = \mathbf{0}$, i.e., if $\mathbf{g}$ is $\mathbb{C}$-differentiable at point $\mathbf{s}$ $(\mathbf{g}(\mathbf{s}_N) = \mathbf{g}(\mathbf{s}) + \mathbf{D}_g(\mathbf{s}_N - \mathbf{s}) + o(\|\mathbf{s}_N - \mathbf{s}\|)$. Finally for real-valued $\boldsymbol{\theta}$, $\mathbf{D}_g^* = \mathbf{D}_{*g}$ and the asymptotic covariance $\mathbf{R}_\theta$ is given by

$$\mathbf{R}_\theta = 2[\mathbf{D}_g \mathbf{R}_s \mathbf{D}_g^H + \mathrm{Re}(\mathbf{D}_g \mathbf{C}_s \mathbf{D}_g^T)]. \tag{7}$$

Under additional regularity assumptions on $\mathbf{g}$, the covariance and the complementary covariance of $\hat{\boldsymbol{\theta}}_N$ are given, respectively, by

$$\mathrm{E}\left[\left(\hat{\boldsymbol{\theta}}_N - \boldsymbol{\theta}\right)(\hat{\boldsymbol{\theta}}_N - \boldsymbol{\theta})^H\right] = \frac{1}{N}\mathbf{R}_\theta + o\left(\frac{1}{N}\right) \quad \text{and}$$
$$\mathrm{E}\left[\left(\hat{\boldsymbol{\theta}}_N - \boldsymbol{\theta}\right)(\hat{\boldsymbol{\theta}}_N - \boldsymbol{\theta})^T\right] = \frac{1}{N}\mathbf{C}_\theta + o\left(\frac{1}{N}\right). \tag{8}$$

Using a second-order expansion of $\mathbf{g}(\mathbf{s}_N)$ where $\mathbf{g}$ is supposed to be $\mathbb{R}$-differentiable to the second-order [8], it is proved in the Appendix that the bias is given by the closed-form expression not published in the open literature:

$$\mathrm{E}\left(\hat{\boldsymbol{\theta}}_N\right) - \boldsymbol{\theta} = \frac{1}{2N}\begin{bmatrix} 2\mathrm{Tr}(\mathbf{R}_s \mathbf{H}_{g,1}^{(1)}) + \mathrm{Tr}(\mathbf{C}_s^* \mathbf{H}_{g,1}^{(2)}) + \mathrm{Tr}(\mathbf{C}_s \mathbf{H}_{g,1}^{(3)}) \\ \vdots \\ 2\mathrm{Tr}(\mathbf{R}_s \mathbf{H}_{g,q}^{(1)}) + \mathrm{Tr}(\mathbf{C}_s^* \mathbf{H}_{g,q}^{(2)}) + \mathrm{Tr}(\mathbf{C}_s \mathbf{H}_{g,q}^{(3)}) \end{bmatrix} + o\left(\frac{1}{N}\right) \tag{9}$$

where $\mathbf{H}_{g,k}^{(1)}$, $\mathbf{H}_{g,k}^{(2)}$ and $\mathbf{H}_{g,k}^{(3)}$ are the three Hessian matrices [5, A2.3] $(\partial/\partial\mathbf{s})(\partial\mathbf{g}_k/\partial\mathbf{s})^H$, $(\partial/\partial\mathbf{s}^*)(\partial\mathbf{g}_k/\partial\mathbf{s})^H$ and $(\partial/\partial\mathbf{s})(\partial\mathbf{g}_k/\partial\mathbf{s})^T$ of the kth component $\mathbf{g}_k$ of the function $\mathbf{g}$ at point $\mathbf{s}$, respectively. We note that for $\mathbf{g}$ $\mathbb{C}$-differentiable to the second-order, the only nonzero Hessian is $\mathbf{H}_{g,k}^{(3)}$, and (9) reduces to its last term which is zero if $\mathbf{s}_N$ is asymptotically proper. So for $\mathbf{g}$ $\mathbb{C}$-differentiable to the second-order and $\mathbf{s}_N$ asymptotically proper, $\hat{\boldsymbol{\theta}}_N$ is asymptotically unbiased to the first order.[5] Finally, for real-valued $\boldsymbol{\theta}$, $\mathbf{H}_{g,k}^{(3)} = (\mathbf{H}_{g,k}^{(2)})^*$, Eq. (9) reduces to

$$\mathrm{E}\left(\hat{\boldsymbol{\theta}}_N\right) - \boldsymbol{\theta} = \frac{1}{2N}\begin{bmatrix} \mathrm{Tr}(\mathbf{R}_{\tilde{s}} \mathbf{H}_{\tilde{g},1}) \\ \vdots \\ \mathrm{Tr}(\mathbf{R}_{\tilde{s}} \mathbf{H}_{\tilde{g},q}) \end{bmatrix} + o\left(\frac{1}{N}\right),$$

where

$$\mathbf{H}_{\tilde{g},k} = \frac{\partial}{\partial\tilde{\mathbf{s}}}\left(\frac{\partial\mathbf{g}}{\partial\tilde{\mathbf{s}}}\right)^H = \begin{bmatrix} \mathbf{H}_{g,k}^{(1)} & \mathbf{H}_{g,k}^{(2)} \\ \mathbf{H}_{g,k^*}^{(2)} & \mathbf{H}_{g,k^*}^{(1)} \end{bmatrix}$$

is the complex augmented Hessian matrix [5, A2.3] of the kth component of the function $\mathbf{g}$ at point $\mathbf{s}$ and

$$\mathbf{R}_{\tilde{s}} = \begin{bmatrix} \mathbf{R}_s & \mathbf{C}_s \\ \mathbf{C}_s^* & \mathbf{R}_s^* \end{bmatrix}$$

is the asymptotic augmented covariance[6] of $\mathbf{s}_N$.

We note that necessary mathematical conditions concerning the remainder terms of these first- and second-order expansions are in the signal processing literature never checked as these conditions are very difficult to prove for the involved mappings $\mathbf{s}_N \overset{\mathbf{g}}{\longmapsto} \hat{\boldsymbol{\theta}}_N$. For example, the following necessary conditions are given in [12, Th. 4.2.2] for the second-order algorithms: (i) the measurements $\{\mathbf{x}(n)\}_{n=1,\ldots,N}$ are independent with finite eighth moments, (ii) the mapping $\mathbf{s}_N \overset{\mathbf{g}}{\longmapsto} \hat{\boldsymbol{\theta}}_N$ is four times $\mathbb{R}$-differentiable, (iii) the fourth derivative of this mapping and those of its square are bounded. These assumptions that do not depend on the distribution of the measurements are very strong, but fortunately (8) and (9) continue to hold in many cases in which these assumptions are not satisfied, in particular for Gaussian distributed data (see, e.g., [12, Ex. 4.2.2]).

Finally, we note that if in practice all functions $\mathbf{g}$, i.e., algorithms are $\mathbb{R}$-differentiable, only some of them are $\mathbb{C}$-differentiable. Among them, when $\hat{\boldsymbol{\theta}}_N$ are roots (e.g., for the root MUSIC algorithms) or explicit solutions (e.g., for the $C(k,q)$ formula [13] extended to the complex case [14]) of polynomials equations whose coefficients are $\mathbb{C}$-differentiable functions of the statistics $\mathbf{s}_N$, the algorithm $\mathbf{g}$ is $\mathbb{C}$-differentiable. This is in contrast to the case where $\hat{\boldsymbol{\theta}}_N$ maximizes a (real-valued) function depending on the statistics $\mathbf{s}_N$, where $\mathbf{g}$ may be now only $\mathbb{R}$-differentiable. This is the case for the subspace-based algorithms for estimating the SIMO and MIMO impulse responses.

3. Asymptotically minimum variance bound

To assess the performance of an algorithm based on a specific statistic $\mathbf{s}_N$ built from $(\mathbf{x}(n))_{n=1,\ldots,N}$, it is interesting to compare the asymptotic covariance $\mathbf{R}_\theta$ (5) and the complementary covariance $\mathbf{C}_\theta$ (5) to an attainable lower bound that depends on the statistic $\mathbf{s}_N$ only. The asymptotically minimum variance bound (AMVB) is such a bound.[7] This bound is generally easy to derive in contrast to the CRB which depends on the distribution of the measurements that appears to be prohibitive to compute for non-Gaussian distributions, except in special cases. This bound uses only the statistical properties of the statistic $\mathbf{s}_N$ and can be used as a benchmark against which potential estimates $\hat{\boldsymbol{\theta}}_N = \mathbf{g}(\mathbf{s}_N)$ are tested. To extend the derivations of Porat and Friedlander [15] concerning this AMVB to complex-valued measurements and parameters, three additional conditions than those introduced in Section 2

[5] This contrasts with the real-valued case for which the bias on $\hat{\boldsymbol{\theta}}_N$ is of order $1/N$.

[6] Note that $\mathbf{R}_{\tilde{s}}$ characterizes the asymptotic second-order moments of $\tilde{\mathbf{s}}_N$ as

$$\mathbf{C}_{\tilde{s}} = \begin{bmatrix} \mathbf{C}_s & \mathbf{R}_s \\ \mathbf{R}_s^* & \mathbf{C}_s^* \end{bmatrix}.$$

[7] Also called asymptotically best consistent (ABC) estimators in [16].

must be satisfied. First, the mapping $\boldsymbol{\theta} \longmapsto \mathbf{s}(\boldsymbol{\theta})$ must be $\mathbb{R}$-differentiable. Second, the involved function $\mathbf{g}$ that defines the considered algorithm must be $\mathbb{R}$-differentiable. And third, the asymptotic augmented covariance $\mathbf{R}_{\tilde{s}}$ of $\mathbf{s}_N$ must be nonsingular. To satisfy this condition, the $2p$ components of $\tilde{\mathbf{s}}_N \stackrel{\text{def}}{=} [\mathbf{s}_N^T, \mathbf{s}_N^H]^T$ must be asymptotically linearly independent random variables. Consequently, no component of $\mathbf{s}_N$ must be real-valued. If some components are real-valued, the redundancies in $\tilde{\mathbf{s}}_N$ must be withdrawn (see, e.g., [17] for the second-order statistics).

Using the augmented representation, and following the steps of the derivation of the AMVB for real-valued $\mathbf{s}_N$ and $\hat{\boldsymbol{\theta}}_N$ [15], it is proved in the Appendix that the augmented covariance matrix $\mathbf{R}_{\tilde{\theta}}$ of the asymptotic distribution of an estimate of $\boldsymbol{\theta}$ given by an arbitrary consistent algorithm (characterized by the mapping $\mathbf{g}$) based on the statistic $\mathbf{s}_N$ is bounded below by $(\tilde{\mathbf{D}}_s^H \mathbf{R}_{\tilde{s}}^{-1} \tilde{\mathbf{D}}_s)^{-1}$:

Result 1.

$$\mathbf{R}_{\tilde{\theta}} = \begin{bmatrix} \mathbf{R}_\theta & \mathbf{C}_\theta \\ \mathbf{C}_\theta^* & \mathbf{R}_\theta^* \end{bmatrix} = \tilde{\mathbf{D}}_g \mathbf{R}_{\tilde{s}} \tilde{\mathbf{D}}_g^H \geq \text{AMVB}_s(\tilde{\boldsymbol{\theta}}) \stackrel{\text{def}}{=} (\tilde{\mathbf{D}}_s^H \mathbf{R}_{\tilde{s}}^{-1} \tilde{\mathbf{D}}_s)^{-1} \tag{10}$$

with $\tilde{\mathbf{D}}_g \stackrel{\text{def}}{=} \begin{bmatrix} \mathbf{D}_g & \mathbf{D}_{*,g} \\ \mathbf{D}_{*,g}^* & \mathbf{D}_g^* \end{bmatrix}$ and $\tilde{\mathbf{D}}_s \stackrel{\text{def}}{=} \begin{bmatrix} \mathbf{D}_s & \mathbf{D}_{*,s} \\ \mathbf{D}_{*,s}^* & \mathbf{D}_s^* \end{bmatrix}$ where $\mathbf{D}_s$ and $\mathbf{D}_{*,s}$ denote the $\mathbb{R}$-derivative and the conjugate $\mathbb{R}$-derivative of $\mathbf{s}(\boldsymbol{\theta})$ at point $\boldsymbol{\theta}$, respectively.

Using the partitioned matrix-inversion lemma in (10), $\mathbf{R}_\theta$ is lower bounded as well. But an algorithm that attains this bound alone does not necessarily attain the AMVB (10) since $\mathbf{R}_\theta$ does not provide a full second-order description of a complex random variable; $\mathbf{C}_\theta$ is also needed.

Furthermore, it is proved in the Appendix that the following nonlinear least squares algorithm is an algorithm that attains the AMVB:

$$\hat{\boldsymbol{\theta}}_N = \arg\min_{\boldsymbol{\beta}} [\tilde{\mathbf{s}}_N - \tilde{\mathbf{s}}(\boldsymbol{\beta})]^H \mathbf{W}_N [\tilde{\mathbf{s}}_N - \tilde{\mathbf{s}}(\boldsymbol{\beta})], \tag{11}$$

where $\tilde{\mathbf{s}}(\boldsymbol{\beta}) \stackrel{\text{def}}{=} [\mathbf{s}^T(\boldsymbol{\beta}), \mathbf{s}^H(\boldsymbol{\beta})]^T$ and $\mathbf{W}_N$ is an arbitrary consistent estimate of $\mathbf{R}_{\tilde{s}}^{-1}$ that satisfies $\mathbf{W}_N = \mathbf{R}_{\tilde{s}}^{-1} + o(\|\mathbf{s}_N - \mathbf{s}(\boldsymbol{\theta})\|)$.

For real-valued $\boldsymbol{\theta}$, $\mathbf{D}_{*,g} = \mathbf{D}_g$ and the AMVB (10) reduces to

$$\mathbf{R}_\theta = 2[\mathbf{D}_g \mathbf{R}_s \mathbf{D}_g^H + \text{Re}(\mathbf{D}_g \mathbf{C}_s \mathbf{D}_g^T)] \geq \text{AMVB}_s(\boldsymbol{\theta}) \stackrel{\text{def}}{=} \left(\left[\mathbf{D}_s^H, \mathbf{D}_s^T \right] \mathbf{R}_{\tilde{s}}^{-1} \begin{bmatrix} \mathbf{D}_s \\ \mathbf{D}_s^* \end{bmatrix} \right)^{-1}.$$

An example of such a derivation is given in [17] for the second-order statistics applied to DOA estimation. Note that in contrast to the Cramer–Rao bound (CRB) that is generally difficult to compute for nonGaussian distributions, the AMVB that uses only the asymptotic second-order statistics of $\mathbf{s}_N$ is much easier to derive.

4. Cramer–Rao bound

To simplify the notations, when the number N of measurements is fixed, these measurements are denoted by $\mathbf{x}$ and their PDF by $p(\mathbf{x};\boldsymbol{\theta})$, where throughout this section, $\boldsymbol{\theta}$ denotes the unknown parameter that gathers the parameter of interest and nuisance parameter.

4.1. General properties of the FIM

Many authors have extended the CRB to complex-valued measurements and parameters. Among them, Ref. [18] has derived this bound by imitating the proof in the real case and Ref. [19] has used the one-to-one mappings $\tilde{\mathbf{x}} \longleftrightarrow \overline{\mathbf{x}}$ and $\tilde{\boldsymbol{\theta}} \longleftrightarrow \overline{\boldsymbol{\theta}}$. Note that despite the CRB has been well covered in the complex case, new contributions continue to appear (see, e.g., [20]).

If $\hat{\boldsymbol{\theta}}$ denotes an unbiased estimator of $\boldsymbol{\theta}$, the augmented covariance matrix of $\hat{\boldsymbol{\theta}}$, $\mathbf{R}_{\tilde{\hat{\theta}}} = \begin{bmatrix} \mathbf{R}_{\hat{\theta}} & \mathbf{C}_{\hat{\theta}} \\ \mathbf{C}_{\hat{\theta}}^* & \mathbf{R}_{\hat{\theta}}^* \end{bmatrix}$, where $\mathbf{R}_{\hat{\theta}} \stackrel{\text{def}}{=} \text{E}[(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^H]$ and $\mathbf{C}_{\hat{\theta}} \stackrel{\text{def}}{=} \text{E}[(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T]$, is upper bounded by the inverse of the augmented Fisher information matrix (FIM):

$$\mathbf{J}_{\tilde{\theta}} = \begin{bmatrix} \mathbf{J}_\theta & \mathbf{J}_{*,\theta} \\ \mathbf{J}_{*,\theta}^* & \mathbf{J}_\theta^* \end{bmatrix} \tag{12}$$

assumed to be nonsingular [19, Theorem 1], [5, Result 6.3]:

$$\mathbf{R}_{\tilde{\hat{\theta}}} \geq \text{CRB}(\tilde{\boldsymbol{\theta}}) \stackrel{\text{def}}{=} \mathbf{J}_{\tilde{\theta}}^{-1}. \tag{13}$$

where $\mathbf{J}_\theta$ and $\mathbf{J}_{*,\theta}$ are the complex FIM and the complementary complex FIM, respectively, given under regularity conditions by

$$\mathbf{J}_\theta = \text{E}\left(\left[\frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right]^H \left[\frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right] \right) = -\text{E}\left(\frac{\partial}{\partial \boldsymbol{\theta}} \left[\frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right]^H \right), \tag{14}$$

$$\mathbf{J}_{*,\theta} = \text{E}\left(\left[\frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right]^H \left[\frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^*} \right] \right) = -\text{E}\left(\frac{\partial}{\partial \boldsymbol{\theta}^*} \left[\frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^*} \right]^T \right). \tag{15}$$

The CRB (13) implies the following bound on the covariance matrix $\mathbf{R}_{\hat{\theta}}$ of $\hat{\boldsymbol{\theta}}$ [18]:

$$\mathbf{R}_{\hat{\theta}} \geq (\mathbf{J}_\theta - \mathbf{J}_{*,\theta} \mathbf{J}_\theta^{-1} \mathbf{J}_{*,\theta}^*)^{-1} \geq \mathbf{J}_\theta^{-1}. \tag{16}$$

If an unbiased estimator $\hat{\boldsymbol{\theta}}$ attains this bound on $\mathbf{R}_{\hat{\theta}}$ alone, it does not imply that $\hat{\boldsymbol{\theta}}$ attains the CRB (13), since also $\mathbf{C}_{\hat{\theta}} = -(\mathbf{J}_\theta - \mathbf{J}_{*,\theta} \mathbf{J}_\theta^{-1} \mathbf{J}_{*,\theta}^*)^{-1} \mathbf{J}_{*,\theta} \mathbf{J}_\theta^{*-1}$ needs to hold (see also [19, Corollary 1(a)]). Only if the complementary FIM $\mathbf{J}_{*,\theta}$ vanishes, then $\mathbf{R}_{\hat{\theta}} = \mathbf{J}_\theta^{-1}$ implies that $\hat{\boldsymbol{\theta}}$ attains the CRB (13).

Note that (16) assumes that $\mathbf{J}_{\tilde{\theta}}$ is nonsingular, which is not the case for real-valued parameters for which $\mathbf{J}_\theta = \mathbf{J}_{*,\theta}$. In this case, the complex CRB is simply given by $\mathbf{R}_{\hat{\theta}} \geq \mathbf{J}_\theta^{-1}$.

In the presence of nuisance parameters $\boldsymbol{\alpha}$ (generally real-valued), the complex CRB on the parameter of interest $\boldsymbol{\theta}$ only is obtained similar to that in the real case. Using the one-to-one mapping $\begin{bmatrix} \overline{\boldsymbol{\theta}} \\ \boldsymbol{\alpha} \end{bmatrix} \longleftrightarrow \begin{bmatrix} \hat{\boldsymbol{\theta}} \\ \boldsymbol{\alpha} \end{bmatrix}$, it is straightforward to prove the following result not published in the open literature:

In the case of nuisance parameter $\boldsymbol{\alpha}$, (13) and (16), respectively, become

$$\mathbf{R}_{\tilde{\hat{\theta}}} \geq \text{CRB}(\tilde{\boldsymbol{\theta}}) \stackrel{\text{def}}{=} (\mathbf{J}_{\tilde{\theta}} - \mathbf{J}_{\tilde{\theta},\alpha} \mathbf{J}_\alpha^{-1} \mathbf{J}_{\tilde{\theta},\alpha}^H)^{-1} \geq \mathbf{J}_{\tilde{\theta}}^{-1}, \tag{17}$$

$$\mathbf{R}_{\hat{\theta}} \geq [(\mathbf{J}_{\tilde{\theta}} - \mathbf{J}_{\tilde{\theta},\alpha} \mathbf{J}_\alpha^{-1} \mathbf{J}_{\tilde{\theta},\alpha}^H)^{-1}]_{\theta,\theta} \geq (\mathbf{J}_\theta - \mathbf{J}_{*,\theta} \mathbf{J}_\theta^{-1} \mathbf{J}_{*,\theta}^*)^{-1} \geq \mathbf{J}_\theta^{-1}, \tag{18}$$

where $[\cdot]_{\theta,\theta}$ denotes the $q \times q$ top-left submatrix of $[\cdot]$, $\mathbf{J}_\alpha$ is the usual FIM w.r.t. the real-valued parameter $\boldsymbol{\alpha}$ only,

and $\mathbf{J}_{\tilde{\theta},\alpha} = \begin{bmatrix} \mathbf{J}_{\theta,\alpha} \\ \mathbf{J}_{*,\theta,\alpha} \end{bmatrix}$ with

$$\mathbf{J}_{\theta,\alpha} = \mathrm{E}\left(\left[\frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta},\boldsymbol{\alpha})}{\partial \boldsymbol{\theta}}\right]^H \left[\frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta},\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}}\right]\right) = -\mathrm{E}\left(\frac{\partial}{\partial \boldsymbol{\alpha}}\left[\frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta},\boldsymbol{\alpha})}{\partial \boldsymbol{\theta}}\right]^H\right) \tag{19}$$

$$\mathbf{J}_{*,\theta,\alpha} = \mathrm{E}\left(\left[\frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta},\boldsymbol{\alpha})}{\partial \boldsymbol{\theta}^*}\right]^H \left[\frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta},\boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}}\right]\right) = -\mathrm{E}\left(\frac{\partial}{\partial \boldsymbol{\alpha}}\left[\frac{\partial \ln p(\mathbf{x};\boldsymbol{\theta},\boldsymbol{\alpha})}{\partial \boldsymbol{\theta}^*}\right]^H\right). \tag{20}$$

4.2. Specific Gaussian case

4.2.1. Slepian–Bangs formula

For complex Gaussian distributions, $\mathcal{N}_C(\mathbf{m}_x;\mathbf{R}_x,\mathbf{C}_x)$, the Slepian–Bangs formula has been extended in [21] and [5, 6.3.5] for real and complex-valued parameters, respectively, where their elementwise FIM and the complementary FIM have been given.[8] Note that these matrices can also be given by the following compact expressions:

$$\mathbf{J}_\theta = \left(\frac{\partial \mathbf{m}_{\tilde{x}}}{\partial \boldsymbol{\theta}}\right)^H \mathbf{R}_{\tilde{x}}^{-1} \frac{\partial \mathbf{m}_{\tilde{x}}}{\partial \boldsymbol{\theta}} + \frac{1}{2}\mathbf{D}_{\mathbf{r}_{\tilde{x}}}^H \mathbf{C}_{\mathbf{r}_{\tilde{x}}}^{-1*} \mathbf{D}_{\mathbf{r}_{\tilde{x}}}, \tag{21}$$

$$\mathbf{J}_{*,\theta} = \left(\frac{\partial \mathbf{m}_{\tilde{x}}}{\partial \boldsymbol{\theta}}\right)^H \mathbf{R}_{\tilde{x}}^{-1} \frac{\partial \mathbf{m}_{\tilde{x}}}{\partial \boldsymbol{\theta}^*} + \frac{1}{2}\mathbf{D}_{\mathbf{r}_{\tilde{x}}}^H \mathbf{C}_{\mathbf{r}_{\tilde{x}}}^{-1*} \mathbf{D}_{*,\mathbf{r}_{\tilde{x}}}, \tag{22}$$

which gives

$$\mathbf{J}_{\tilde{\theta}} = \begin{bmatrix} \left(\frac{\partial \mathbf{m}_{\tilde{x}}}{\partial \theta}\right)^H \\ \left(\frac{\partial \mathbf{m}_{\tilde{x}}}{\partial \theta^*}\right)^H \end{bmatrix} \mathbf{R}_{\tilde{x}}^{-1} \begin{bmatrix} \frac{\partial \mathbf{m}_{\tilde{x}}}{\partial \boldsymbol{\theta}} & \frac{\partial \mathbf{m}_{\tilde{x}}}{\partial \boldsymbol{\theta}^*} \end{bmatrix} + \frac{1}{2}\begin{bmatrix} \mathbf{D}_{\mathbf{r}_{\tilde{x}}}^H \\ \mathbf{D}_{*,\mathbf{r}_{\tilde{x}}}^H \end{bmatrix} \mathbf{R}_{\mathbf{r}_{\tilde{x}}}^{-1*} \begin{bmatrix} \mathbf{D}_{\mathbf{r}_{\tilde{x}}} & \mathbf{D}_{*,\mathbf{r}_{\tilde{x}}} \end{bmatrix}, \tag{23}$$

where $\mathbf{m}_{\tilde{x}} = (\mathbf{m}_x^T, \mathbf{m}_x^H)^T$, $\mathbf{R}_{\tilde{x}} = \mathrm{E}((\tilde{\mathbf{x}} - \mathbf{m}_{\tilde{x}})(\tilde{\mathbf{x}} - \mathbf{m}_{\tilde{x}})^H)$ with $\tilde{\mathbf{x}} = (\mathbf{x}^T, \mathbf{x}^H)^T$, $\mathbf{D}_{\mathbf{r}_{\tilde{x}}}$ and $\mathbf{D}_{*,\mathbf{r}_{\tilde{x}}}$ denote the $\mathbb{R}$-derivative $\partial \mathbf{r}_{\tilde{x}}/\partial \boldsymbol{\theta}$ and the conjugate $\mathbb{R}$-derivative $\partial \mathbf{r}_{\tilde{x}}/\partial \boldsymbol{\theta}^*$ of $\mathbf{r}_{\tilde{x}} \stackrel{\mathrm{def}}{=} \mathrm{vec}(\mathbf{R}_{\tilde{x}})$, respectively, and where $\mathbf{R}_{\mathbf{r}_{\tilde{x}}} = \mathbf{R}_{\tilde{x}}^* \otimes \mathbf{R}_{\tilde{x}}$ is the covariance of the asymptotic distribution of $\mathbf{r}_{\tilde{x},N} = \mathrm{vec}(\mathbf{R}_{\tilde{x},N})$ with $\mathbf{R}_{\tilde{x},N} \stackrel{\mathrm{def}}{=} (1/N)\sum_{n=1}^N \tilde{\mathbf{x}}(n)\tilde{\mathbf{x}}^H(n)$.

4.2.2. Whittle formula

When $\mathbf{x}(n)$ is a real-valued stationary zero-mean Gaussian multivariate process with spectrum $\mathbf{R}_x(f)$ that depends on the real-valued parameter $\boldsymbol{\theta}$, the Whittle formula [22, Th. 9] gives the elements of the asymptotic FIM associated with N sample values of $\mathbf{x}(n)$. Thus the matrix-valued Cramer–Rao bound is given by

$$\mathbf{R}_{\hat{\theta}} \geq \mathbf{J}_\theta^{-1} \quad \text{where } \mathbf{J}_\theta = \frac{N}{2}\int_{-1/2}^{+1/2} \mathbf{D}_{\mathbf{r}_x}^H(f)\left(\mathbf{R}_x^{-*}(f) \otimes \mathbf{R}_x^{-1}(f)\right)\mathbf{D}_{\mathbf{r}_x}(f)\, df, \tag{24}$$

where $\mathbf{R}_{\hat{\theta}}$ is the covariance of any unbiased estimate of $\boldsymbol{\theta}$ built from $(\mathbf{x}(n))_{n=1,\ldots,N}$ and $\mathbf{D}_{\mathbf{r}_x}(f)$ denotes the derivative $\partial \mathbf{r}_x(f)/\partial \boldsymbol{\theta}$ of $\mathbf{r}_x(f) \stackrel{\mathrm{def}}{=} \mathrm{vec}(\mathbf{R}_x(f))$ where $\mathbf{R}_x(f)$ is Hermitian structured.

Using the one-to-one mappings $\tilde{\mathbf{x}}(n) \longleftrightarrow \overline{\mathbf{x}}(n)$ and $\tilde{\boldsymbol{\theta}} \longleftrightarrow \overline{\boldsymbol{\theta}}$, it is proved in the Appendix the following extension of the Whittle formula:

Result 2. Let $\mathbf{x}(n)$ be a complex-valued stationary zero-mean non necessarily circular, Gaussian multivariate process with spectrum $\mathbf{R}_x(f)$ and complementary spectrum $\mathbf{C}_x(f)$ [5, Section 8.1] that both depend on the complex-valued parameter $\boldsymbol{\theta}$, the matrix-valued Cramer–Rao bound is given by

$$\mathbf{R}_{\hat{\theta}} \geq \mathbf{J}_{\tilde{\theta}}^{-1} \quad \text{with } \mathbf{J}_{\tilde{\theta}} = \begin{bmatrix} \mathbf{J}_\theta & \mathbf{J}_{*,\theta} \\ \mathbf{J}_{*,\theta}^* & \mathbf{J}_\theta^* \end{bmatrix} \quad \text{is assumed to be nonsingular,} \tag{25}$$

and $\mathbf{J}_\theta$ and $\mathbf{J}_{*,\theta}$ are given, respectively, by

$$\mathbf{J}_\theta = \frac{N}{2}\int_{-1/2}^{+1/2} \mathbf{D}_{\mathbf{r}_{\tilde{x}}}^H(f)\left(\mathbf{R}_{\tilde{x}}^{-1*}(f) \otimes \mathbf{R}_{\tilde{x}}^{-1}(f)\right)\mathbf{D}_{\mathbf{r}_{\tilde{x}}}(f)\, df, \tag{26}$$

$$\mathbf{J}_{*,\theta} = \frac{N}{2}\int_{-1/2}^{+1/2} \mathbf{D}_{\mathbf{r}_{\tilde{x}}}^H(f)\left(\mathbf{R}_{\tilde{x}}^{-1*}(f) \otimes \mathbf{R}_{\tilde{x}}^{-1}(f)\right)\mathbf{D}_{*,\mathbf{r}_{\tilde{x}}}(f)\, df, \tag{27}$$

or more compactly, by

$$\mathbf{J}_{\tilde{\theta}} = \frac{N}{2}\int_{-1/2}^{+1/2} \begin{bmatrix} \mathbf{D}_{\mathbf{r}_{\tilde{x}}}^H(f) \\ \mathbf{D}_{\mathbf{r}_{*,\tilde{x}}}^H(f) \end{bmatrix} \mathbf{R}_{\tilde{x}}^{-*}(f) \otimes \mathbf{R}_{\tilde{x}}^{-1}(f)\begin{bmatrix} \mathbf{D}_{\mathbf{r}_{\tilde{x}}}(f) & \mathbf{D}_{\mathbf{r}_{*,\tilde{x}}}(f) \end{bmatrix} df, \tag{28}$$

where $\mathbf{D}_{\mathbf{r}_{\tilde{x}}}(f)$ and $\mathbf{D}_{*,\mathbf{r}_{\tilde{x}}}(f)$ denote the $\mathbb{R}$-derivative $\partial \mathbf{r}_{\tilde{x}}(f)/\partial \boldsymbol{\theta}$ and the conjugate $\mathbb{R}$-derivative $\partial \mathbf{r}_{\tilde{x}}(f)/\partial \boldsymbol{\theta}^*$ of $\mathbf{r}_{\tilde{x}}(f) \stackrel{\mathrm{def}}{=} \mathrm{vec}(\mathbf{R}_{\tilde{x}}(f))$, respectively, and $\mathbf{R}_{\tilde{x}}(f)$ is the spectrum of the augmented process $\tilde{\mathbf{x}}(n)$:

$$\mathbf{R}_{\tilde{x}}(f) = \begin{bmatrix} \mathbf{R}_x(f) & \mathbf{C}_x(f) \\ \mathbf{C}_x^*(-f) & \mathbf{R}_x^*(-f) \end{bmatrix},$$

with $\mathbf{R}_x(f)$ and $\mathbf{C}_x(f)$ the Fourier transforms of $\mathbf{R}_x(k) = \mathrm{E}[(\mathbf{x}(n)\mathbf{x}^H(n-k)]$ and $\mathbf{C}_x(k) = \mathrm{E}[(\mathbf{x}(n)\mathbf{x}^T(n-k)]$, respectively, both characterizing the statistical properties of the random process $\mathbf{x}(n)$.

Note that for real-valued parameters, (25) reduces to $\mathbf{R}_{\hat{\theta}} \geq \mathbf{J}_\theta^{-1}$ that was proved in [23] for deriving the CRB of estimated delays in the context of complex-valued stationary processes.

4.2.3. Circular to noncircular comparison

For the Gaussian distribution characterized by $(\mathbf{m}_x, \mathbf{R}_x, \mathbf{C}_x)$, suppose now that the parameter $\boldsymbol{\theta}$ is identifiable from $(\mathbf{m}_x, \mathbf{R}_x)$ only. A question remains open (see [5, Section 6.3.5]): Is $\mathbf{J}_{\tilde{\theta}}$ more positive definite if $\mathbf{C}_x \neq \mathbf{0}$ than if $\mathbf{C}_x = \mathbf{0}$? Or in other words, does the noncircular case generally improve the CRB of $\boldsymbol{\theta}$ with respect to the circular case? Addressing generally this question from (21) and (22) seems very challenging. But formulating this question in the framework of measurements $\mathbf{x} = (\mathbf{x}(n))_{n=1,\ldots,N}$, where $\mathbf{x}(n)$ are IID and where $\mathbf{m}_x$, $\mathbf{R}_x$ and $\mathbf{C}_x$ denote the mean, covariance and complementary covariance of $\mathbf{x}(n)$, respectively, is much easier, as the AMVB based on the statistics that include both the sample mean, sample covariance and sample complementary covariance attains the CRB for the Gaussian distribution, i.e.,

$$\mathbf{J}_{\tilde{\theta}}^{-1} = (\tilde{\mathbf{D}}_s^H \mathbf{R}_{\tilde{s}}^{-1} \tilde{\mathbf{D}}_s)^{-1}, \tag{29}$$

where $\mathbf{J}_{\tilde{\theta}}$ is associated with $\mathbf{x}(n)$ alone and where the augmented statistics involved is $\tilde{\mathbf{s}}(\boldsymbol{\theta}) = [\mathbf{m}_x^T, \mathbf{m}_x^H, \mathrm{vec}^T(\mathbf{R}_x)$,

[8] Note that (21) and (22) are slightly different from the elementwise expressions [5, (6.65)] and [5, (6.66)] because the latter expressions are erroneous.

$\mathrm{v}^T(\mathbf{C}_x), \mathrm{v}^H(\mathbf{C}_x)]^T$ in order to satisfy the three conditions of the AMVB (10). Using (29), it is proved in the Appendix:

Result 3. When the parameter $\boldsymbol{\theta}$ of the Gaussian distribution (characterized by $(\mathbf{m}_x, \mathbf{R}_x, \mathbf{C}_x)$) of $\mathbf{x}(n)$ is identifiable from $(\mathbf{m}_x, \mathbf{R}_x)$ only, noncircular Gaussian distributions generally improve the CRB for $\boldsymbol{\theta}$ with respect to the circular case:

$$\mathrm{CRB}_{\mathbf{m}_x,\mathbf{R}_x,\mathbf{C}_x}(\tilde{\boldsymbol{\theta}}) \le \mathrm{CRB}_{\mathbf{m}_x,\mathbf{R}_x,\mathbf{0}}(\tilde{\boldsymbol{\theta}}) \tag{30}$$

where $\mathrm{CRB}_{\mathbf{m}_x,\mathbf{R}_x,\mathbf{C}_x}(\tilde{\boldsymbol{\theta}})$ and $\mathrm{CRB}_{\mathbf{m}_x,\mathbf{R}_x,\mathbf{0}}(\tilde{\boldsymbol{\theta}})$ denote the augmented complex CRB (13) associated with noncircular and circular Gaussian distribution, respectively.

In the presence of nuisance parameters $\boldsymbol{\alpha}$, the previous question is much more involved because the complementary covariance matrix $\mathbf{C}_x$ can not only bear information on the parameter of interest $\boldsymbol{\theta}$, but can also introduce additional nuisance parameters. An example in which (30) is not satisfied in the presence of nuisance parameters is presented in Section 6. However in particular statistical models, (30) can be extended as it is proved in the next section.

5. Noisy linear mixture

Consider the following model:

$$\mathbf{x}(n) = \mathbf{A}(\boldsymbol{\theta})\mathbf{s}(n) + \mathbf{e}(n) \in \mathbb{C}^r \quad n = 1, \ldots, N, \tag{31}$$

where (i) $\mathbf{s}(n)$ and $\mathbf{e}(n)$ are independent zero-mean random variables, (ii) $\mathbf{e}(n)$ is circular with $\mathrm{E}(\mathbf{e}(n)\mathbf{e}^H(n)) = \sigma_e^2\mathbf{I}$ and $\mathbf{s}(n) \in \mathbb{C}^p$ is either circular with $\mathrm{E}(\mathbf{s}(n)\mathbf{s}^H(n)) = \mathbf{R}_s$ nonsingular or noncircular with $\mathrm{E}(\tilde{\mathbf{s}}(n)\tilde{\mathbf{s}}^H(n)) = \mathbf{R}_{\tilde{s}}$ nonsingular, (iii) the useful parameter $\boldsymbol{\theta} \in \mathbb{C}^q$ is characterized by the subspace generated by the columns of the full column rank $r \times p$ matrix $\mathbf{A}(\boldsymbol{\theta})$ with $p < r$. The nuisance parameters $\boldsymbol{\alpha}$ gather here the terms $(\mathbf{R}_s)_{i,j}$ for $1 \le i \le j \le p$ and σ_e^2 [resp., the terms $(\mathbf{R}_s)_{i,j}$ and $(\mathbf{C}_s)_{i,j}$ for $1 \le i \le j \le p$ and σ_e^2] in the circular [resp., noncircular] case.

5.1. CRB expressions

Using the direct approach introduced by [24] to concentrate the CRB on the parameter $\boldsymbol{\theta}$, it is proved in the Appendix the following result not been published in the open literature:

For the model (31) with assumptions (i)–(iii) where $(\mathbf{s}(n))_{n=1,\ldots,N}$ and $(\mathbf{e}(n))_{n=1,\ldots,N}$ are independent Gaussian distributed random variables, the CRB for the real-valued parameter alone $\overline{\boldsymbol{\theta}} \stackrel{\mathrm{def}}{=} [\mathrm{Re}^T(\boldsymbol{\theta}), \mathrm{Im}^T(\boldsymbol{\theta})]^T$ is given by

$$\mathrm{CRB}(\overline{\boldsymbol{\theta}}) = \frac{\sigma_e^2}{2N}\left[\mathrm{Re}\left(\frac{\partial \mathbf{a}^H}{\partial \overline{\boldsymbol{\theta}}}\left(\mathbf{H}^T \otimes \boldsymbol{\Pi}_A^{\perp}\right)\frac{\partial \mathbf{a}}{\partial \overline{\boldsymbol{\theta}}}\right)\right]^{-1}, \tag{32}$$

where $\mathbf{a} \stackrel{\mathrm{def}}{=} \mathrm{vec}(\mathbf{A})$, $\boldsymbol{\Pi}_A^{\perp} \stackrel{\mathrm{def}}{=} \mathbf{I} - \mathbf{A}(\mathbf{A}^H\mathbf{A})^{-1}\mathbf{A}^H$ is the orthocomplement of the projection matrix on the columns of $\mathbf{A}$ and $\mathbf{H}$ is given by the Hermitian matrices $\mathbf{R}_s\mathbf{A}^H\mathbf{R}_x^{-1}\mathbf{A}\mathbf{R}_s$ and $\left[\mathbf{R}_s\mathbf{A}^H, \mathbf{C}_s\mathbf{A}^T\right]\mathbf{R}_{\tilde{x}}^{-1}\left[\begin{smallmatrix}\mathbf{A}\mathbf{R}_s\\ \mathbf{A}^*\mathbf{C}_s^*\end{smallmatrix}\right]$ in the circular and noncircular cases, respectively.

We note that (32) extends the CRB compact expression [24, rel. (5)] given for the DOA modeling with scalar-sensors for one parameter per source, and encompasses DOA modeling with vector-sensors for an arbitrary number of parameters per source and many other models as the SIMO and MIMO modelings.

Using the one-to-one mapping $\overline{\boldsymbol{\theta}} \longleftrightarrow \tilde{\boldsymbol{\theta}}$, the following compact expression of the augmented complex CRB (12) and (13) is proved in the Appendix.

Result 4. For the model (31) with assumptions (i)–(iii), where $(\mathbf{s}(n))_{n=1,\ldots,N}$ and $(\mathbf{e}(n))_{n=1,\ldots,N}$ are independent Gaussian distributed random variables, we have

$$\mathbf{R}_{\tilde{\theta}} \ge \mathrm{CRB}(\tilde{\boldsymbol{\theta}}), \quad \text{with} \quad \mathrm{CRB}(\tilde{\boldsymbol{\theta}}) = \mathbf{J}_{\tilde{\theta}}^{-1} = \begin{bmatrix} \mathbf{J}_\theta & \mathbf{J}_{*,\theta} \\ \mathbf{J}_{*,\theta}^* & \mathbf{J}_\theta^* \end{bmatrix}^{-1}, \tag{33}$$

where

$$\mathbf{J}_\theta = \frac{N}{\sigma_e^2}\left[\left(\frac{\partial \mathbf{a}}{\partial \boldsymbol{\theta}}\right)^H\left(\mathbf{H}^T \otimes \boldsymbol{\Pi}_A^{\perp}\right)\left(\frac{\partial \mathbf{a}}{\partial \boldsymbol{\theta}}\right) + \left(\frac{\partial \mathbf{a}}{\partial \boldsymbol{\theta}^*}\right)^T\left(\mathbf{H} \otimes \boldsymbol{\Pi}_A^{\perp T}\right)\left(\frac{\partial \mathbf{a}}{\partial \boldsymbol{\theta}^*}\right)^*\right], \tag{34}$$

$$\mathbf{J}_{*,\theta} = \frac{N}{\sigma_e^2}\left[\left(\frac{\partial \mathbf{a}}{\partial \boldsymbol{\theta}}\right)^H\left(\mathbf{H}^T \otimes \boldsymbol{\Pi}_A^{\perp}\right)\left(\frac{\partial \mathbf{a}}{\partial \boldsymbol{\theta}^*}\right) + \left(\frac{\partial \mathbf{a}}{\partial \boldsymbol{\theta}^*}\right)^T\left(\mathbf{H} \otimes \boldsymbol{\Pi}_A^{\perp T}\right)\left(\frac{\partial \mathbf{a}}{\partial \boldsymbol{\theta}}\right)^*\right]. \tag{35}$$

In the particular case where $\mathbf{a}$ is $\mathbb{C}$-differentiable w.r.t. $\boldsymbol{\theta}$ (e.g., for SIMO and MIMO channel modelings), $\partial\mathbf{a}/\partial\boldsymbol{\theta}^* = \mathbf{0}$, and (34) and (35) reduce to

$$\mathbf{J}_\theta = \frac{N}{\sigma_e^2}\left(\frac{\partial \mathbf{a}}{\partial \boldsymbol{\theta}}\right)^H\left(\mathbf{H}^T \otimes \boldsymbol{\Pi}_A^{\perp}\right)\left(\frac{\partial \mathbf{a}}{\partial \boldsymbol{\theta}}\right) \quad \text{and} \quad \mathbf{J}_{*,\theta} = \mathbf{0}, \tag{36}$$

and the AMV estimator is asymptotically circular with $\mathbf{R}_{\hat{\theta}} = \mathbf{J}_\theta^{-1}$, whatever the circularity properties of $\mathbf{x}(n)$.

Note that the closed-form expressions (32) and (34)–(36) do not take into account the prior knowledge relative to the sources because they have been derived without any constraint on $\mathbf{R}_s$ and $\mathbf{C}_s$. But unfortunately, taking into account these constraints leads to very intricate expressions (see, e.g., [25, Eq. (13)] for circular uncorrelated sources for the DOA modeling). This point will be illustrated in Section 6 with the SISO channel modeling. Furthermore, note that the condition $\mathbf{A}$ is full column rank with $p < r$ which is not necessary to identify the useful parameter $\boldsymbol{\theta}$ when specific a priori knowledge about the sources is available, see, e.g., [28] for real-valued or QPSK modulations and [29] for offset linear modulations in SISO channel modeling.

Finally comparing the circular to the noncircular cases, it is proved in the Appendix that the CRB for $\boldsymbol{\theta}$ in the noncircular case is upper bounded by the associated asymptotic RB in the circular case. More precisely, for the model (31) with assumptions (i)–(iii), where $(\mathbf{s}(n))_{n=1,\ldots,N}$ and $(\mathbf{e}(n))_{n=1,\ldots,N}$ are independent Gaussian distributed random variables, we have

$$\mathrm{CRB}_{\mathbf{R}_x,\mathbf{C}_x}(\tilde{\boldsymbol{\theta}}) \le \mathrm{CRB}_{\mathbf{R}_x,\mathbf{0}}(\tilde{\boldsymbol{\theta}}). \tag{37}$$

This result extends the CRB inequality proved in [21] for the DOA parameters. Consequently, when the precision on the parameter $\boldsymbol{\theta}$ is important, it is preferable to use noncircular $\mathbf{s}(k)$ signals (e.g., real-valued) than circular ones, for example for blind SISO, SIMO and MIMO channels identification.

5.2. Efficiency of subspace-based estimators

For the model (31) with assumptions (i)–(iii), many algorithms are consistent subspace-based, i.e., the estimates $\hat{\boldsymbol{\theta}}$ are obtained by exploiting the orthogonality between a sample subspace and a parameter-dependent subspace [26]. In other words, for circular $\mathbf{x}(n)$, these algorithms satisfy the mapping (1) where the statistic $\mathbf{s}_N$ is usually the orthogonal projector $\mathbf{\Pi}_{R_x,N}$ on noise (or signal) associated with the sample covariance $\mathbf{R}_{x,N}=(1/N)\sum_{n=1}^{N}\mathbf{x}(n)\mathbf{x}^H(n)$. To exploit the potential noncircularity of $\mathbf{x}(n)$, the orthogonal projector $\mathbf{\Pi}_{R_{\tilde{x}},N}$ associated with the sample augmented covariance $\mathbf{R}_{\tilde{x},N}=(1/N)\sum_{n=1}^{N}\tilde{\mathbf{x}}(n)\tilde{\mathbf{x}}^H(n)$ or the couple $(\mathbf{\Pi}_{R_x,N},\mathbf{\Pi}_{C_x,N})$ of orthogonal projectors (where $\mathbf{\Pi}_{C_x,N}$ is the orthogonal projector associated with $\mathbf{C}_{x,N}=(1/N)\sum_{n=1}^{N}\mathbf{x}(n)\mathbf{x}^T(n)$) is used. Although, the asymptotic covariance $\mathbf{R}_s$ of the statistics $\mathrm{vec}(\mathbf{\Pi}_{R_x,N})$, $\mathrm{vec}(\mathbf{\Pi}_{R_{\tilde{x}},N})$ and $\mathrm{vec}(\mathbf{\Pi}_{R_x,N},\mathbf{\Pi}_{C_x,N})$ are singular, and thus do not satisfy the third condition introduced in the beginning of Section 3, the following result (not published in the open literature) is proved in the Appendix:

For the model (31) with assumptions (i)–(iii), the AMVB (10) becomes

$$\mathbf{R}_{\tilde{\theta}}=\begin{bmatrix}\mathbf{R}_\theta & \mathbf{C}_\theta\\ \mathbf{C}_\theta^* & \mathbf{R}_\theta^*\end{bmatrix}=\tilde{\mathbf{D}}_g\mathbf{R}_{\tilde{s}}\tilde{\mathbf{D}}_g^H\geq \mathrm{AMVB}_s(\tilde{\boldsymbol{\theta}})\stackrel{\mathrm{def}}{=}(\tilde{\mathbf{D}}_s^H\mathbf{R}_{\tilde{s}}^{\#}\tilde{\mathbf{D}}_s)^{-1} \tag{38}$$

for $\mathbf{s}_N=\mathrm{vec}(\mathbf{\Pi}_{R_x,N})$, $\mathrm{vec}(\mathbf{\Pi}_{R_{\tilde{x}},N})$ or $\mathrm{vec}(\mathbf{\Pi}_{R_x,N},\mathbf{\Pi}_{C_x,N})$. Furthermore, despite the lack of a one-to-one mapping between $(\mathbf{\Pi}_{R_x,N},\mathbf{\Pi}_{C_x,N})$ and $\mathbf{\Pi}_{R_{\tilde{x}},N}$, contrary to the one-to-one mapping $(\mathbf{R}_{x,N},\mathbf{C}_{x,N})\longleftrightarrow\mathbf{R}_{\tilde{x},N}$, the AMVB based on the statistics $(\mathbf{\Pi}_{R_x,N},\mathbf{\Pi}_{C_x,N})$ and $\mathbf{\Pi}_{R_{\tilde{x}},N}$ coincide. Note that the expression of $\mathbf{R}_{\tilde{s}}$ does not depend on the temporal covariance and the fourth-order moments of $\mathbf{x}(n)$ [21]. So, the asymptotic augmented covariance $\mathbf{R}_{\tilde{\theta}}$ of an estimator of $\boldsymbol{\theta}$ given by an arbitrary consistent subspace-based algorithm built from $\mathbf{\Pi}_{R_x,N}$, $\mathbf{\Pi}_{R_{\tilde{x}},N}$ or $(\mathbf{\Pi}_{R_x,N},\mathbf{\Pi}_{C_x,N})$ depends on the distribution of the time series $\mathbf{x}(n)$ through the second-order moments of $\mathbf{x}(n)$ only.

To evaluate the efficiency of these subspace-based algorithms, we consider now the particular case where $(\mathbf{s}(n))_{n=1,\ldots,N}$ and $(\mathbf{e}(n))_{n=1,\ldots,N}$ are independent Gaussian distributed random variables. In this case, the following result is proved in the Appendix:

Result 5. For the model (31) with assumptions (i)–(iii), where $(\mathbf{s}(n))_{n=1,\ldots,N}$ and $(\mathbf{e}(n))_{n=1,\ldots,N}$ are independent Gaussian distributed random variables, the AMVB (38) associated with the statistics $\mathbf{\Pi}_{R_x,N}$ [resp. $\mathbf{\Pi}_{R_{\tilde{x}},N}$ or $(\mathbf{\Pi}_{R_x,N},\mathbf{\Pi}_{C_x,N})$] are equal to the normalized (with $N=1$) CRB (33) associated with the circular [resp. noncircular] Gaussian distribution of $\mathbf{x}(n)$:

$$\mathrm{AMVB}_s(\tilde{\boldsymbol{\theta}})=\mathrm{CRB}(\tilde{\boldsymbol{\theta}})\quad(\text{with } N=1). \tag{39}$$

This result extends to arbitrary complex parametrization, a result proved in [27] in the particular case of DOA modeling with a single parameter per source. It proves the interest of the subspace-based algorithms when no a priori information is available on the distribution of the signals $\mathbf{s}(n)$ and $\mathbf{e}(n)$.

Finally, using a whitening approach, the following remark is proved in the Appendix:

Remark 1. All the Results of this section (rel. (32), Result 4, rel. (37) and (38), and Result 5) can be extended to the case where the noise $\mathbf{e}(n)$ is circular with $\mathrm{E}(\mathbf{e}(n)\mathbf{e}^H(n))=\sigma_e^2\mathbf{\Sigma}$ where $\mathbf{\Sigma}$ is known positive definite, by replacing $\mathbf{\Pi}_A^{\perp}$ by $\mathbf{\Pi}_\Sigma\stackrel{\mathrm{def}}{=}\mathbf{\Sigma}^{-1}-\mathbf{\Sigma}^{-1}\mathbf{A}(\mathbf{A}^H\mathbf{\Sigma}^{-1}\mathbf{A})^{-1}\mathbf{A}^H\mathbf{\Sigma}^{-1}$, which is no longer a projection matrix.

6. Numerical illustration

In this section, we illustrate the results of Section 5 by considering the complex blind SIMO channel identification and complex independent component analysis (ICA) models. The blind SIMO channel identification data model can be written as shown in (31) after collecting the $L+1$ successive received sampled complex baseband signals at the output of a $1\times P$ SIMO FIR channel of order M where

$$\mathbf{A}(\boldsymbol{\theta})=\begin{pmatrix} h_0^1 & h_1^1 & \cdots & h_M^1 & & \\ \vdots & \vdots & \vdots & \vdots & & \\ h_0^P & h_1^P & \cdots & h_M^P & & \\ & h_0^1 & h_1^1 & \cdots & h_M^1 & \\ & \vdots & \vdots & \vdots & \vdots & \\ & & \vdots & \vdots & \vdots & \vdots \\ & & h_0^P & h_1^P & \cdots & h_M^P \end{pmatrix}\in\mathbb{C}^{P(L+1)\times(L+M+1)},\quad \text{with } {}^{9}\ h_0^1=1$$

with $\boldsymbol{\theta}=[h_0^2,\ldots,h_0^P,h_1^1,\ldots,h_1^P,\ldots,h_M^1,\ldots,h_M^P]^T\in\mathbb{C}^{(M+1)P-1}$ and where $\mathbf{s}(n)$ gathers $L+M+1$ successive inputs. To satisfy the condition (iii) introduced in the beginning of Section 5, L must satisfy $P(L+1)>L+M+1$ and the polynomials $h^{(p)}(z)=\sum_{k=0}^{M}h_k^p z^k$, $p=1,\ldots,P$ must not share common zeros.

We consider here the particular case $P=L=M=2$, where the input $\mathbf{s}(n)$ is a sequence of equiprobable independent BPSK $\sigma_s e^{i\phi_s}\{-1,+1\}$ or QPSK $\sigma_s e^{i\phi_s}\{-1,+1,-i,+i\}$ symbols. Consequently, $\mathbf{R}_s=\sigma_s^2\mathbf{I}$ for both inputs, but $\mathbf{C}_s=\sigma_s^2e^{2i\phi_s}\mathbf{I}$ for BPSK symbols and $\mathbf{C}_s=\mathbf{0}$ for QPSK symbols.

Fig. 1 exhibits the normalized ($N=1$) asymptotic MSE ($\boldsymbol{\theta}$): $\mathrm{Tr}[\mathrm{AMVB}_s^{\mathrm{BPSK}}(\boldsymbol{\theta})]$ and $\mathrm{Tr}[\mathrm{AMVB}_s^{\mathrm{QPSK}}(\boldsymbol{\theta})]$ associated with the projector statistics, as a function of the phase β for the channels $h^{(1)}(z)=(1-z_{1,1}^{-1}z)(1-z_{2,1}^{-1}z)$ and $h^{(2)}(z)=(1-z_{1,2}^{-1}z)(1-z_{2,2}^{-1}z)$ with $z_{1,1}=0.8$, $z_{2,1}=1.25e^{i\pi/4}$, $z_{1,2}=0.8e^{i\beta}$ and $z_{2,2}=1.25e^{-i\pi/4}$, where $\sigma_s^2/\sigma_e^2=10$ dB and $\phi_s=\pi/3$. We note that here the AMV estimators are asymptotically circular because $\mathbf{A}(\boldsymbol{\theta})$ is $\mathbb{C}$-differentiable. For Gaussian distributed inputs $\mathbf{s}(n)$, these AMVBs are equal to the associated CRBs (33) and (36) from Result 2. Fig. 1 shows that the difference between $\mathrm{Tr}[\mathrm{AMVB}_s^{\mathrm{BPSK}}(\boldsymbol{\theta})]$ and $\mathrm{Tr}[\mathrm{AMVB}_s^{\mathrm{QPSK}}(\boldsymbol{\theta})]$ is large enough, in particular for β approaching 0 for which $\mathbf{A}(\boldsymbol{\theta})$ is close to be singular where $\boldsymbol{\theta}$ is not identifiable. This behavior is similar to the DOA modeling for which the difference between these two AMVBs is more prominent for low DOA separations [21].

When the structure information of $\mathbf{R}_s$ and $\mathbf{C}_s$ is used, two new AMVBs ($\mathrm{AMVB}_{r,c}^{\mathrm{BPSK}}(\boldsymbol{\theta})$ and $\mathrm{AMVB}_{r,c}^{\mathrm{QPSK}}(\boldsymbol{\theta})$) based on the

[9] We have assumed $h_0^1=1$ to avoid any ambiguity for the definition of the impulse response $\boldsymbol{\theta}$ in the product $\mathbf{A}(\boldsymbol{\theta})\mathbf{s}(n)$.

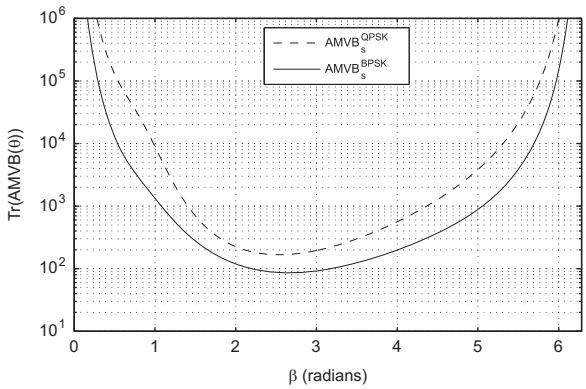

Fig. 1. Tr[AMVB$_s^{BPSK}(\boldsymbol{\theta})$] and Tr[AMVB$_s^{QPSK}(\boldsymbol{\theta})$] as a function of the phase parameter β for the channel.

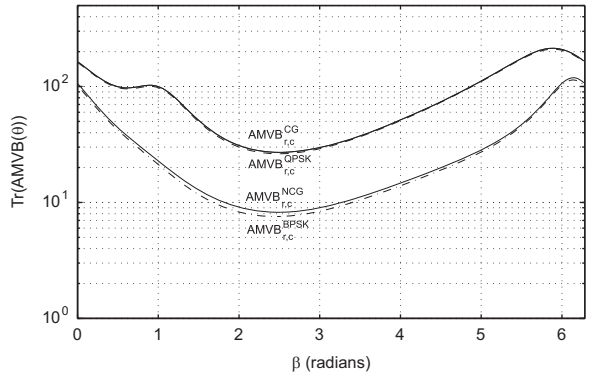

Fig. 2. Tr[AMVB$_{r,c}^{BPSK}(\boldsymbol{\theta})$], Tr[AMVB$_{r,c}^{QPSK}(\boldsymbol{\theta})$], Tr[AMVB$_{r,c}^{CG}(\boldsymbol{\theta})$] and Tr[AMVB$_{r,c}^{NCG}(\boldsymbol{\theta})$] as a function of the phase parameter β for the channel.

statistics $\mathbf{R}_{x,N}=(1/N)\sum_{n=1}^{N}\mathbf{x}(n)\mathbf{x}^H(n)$ and $\mathbf{C}_{x,N}=(1/N)\sum_{n=1}^{N}\mathbf{x}(n)\mathbf{x}^T(n)$ can be considered. Fig. 2 exhibits the normalized ($N=1$) asymptotic MSE ($\boldsymbol{\theta}$): Tr[AMVB$_{r,c}^{BPSK}(\boldsymbol{\theta})$] and Tr[AMVB$_{r,c}^{QPSK}(\boldsymbol{\theta})$] in the same scenario as Fig. 1. The AMBV AMVB$_{r,c}^{NCG}(\boldsymbol{\theta})$ and AMVB$_{r,c}^{CG}(\boldsymbol{\theta})$ associated with noncircular and circular Gaussian distributions, respectively, are also exhibited. This figure shows that these AMVB are slightly sensitive to the distribution of the inputs. Furthermore, the AMVB associated with BPSK and noncircular Gaussian distributed inputs are upper bounded by the AMVB associated with QPSK and circular Gaussian distributed inputs, respectively, despite the presence of the nuisance parameters $\boldsymbol{\alpha}=[\phi_s,\sigma_s,\sigma_e]^T$. Finally, comparing Figs 1 and 2 shows that this uncorrelation a priori information on the inputs is quite informative. Moreover, we see that these bounds keep finite values when $\mathbf{A}$ is no longer full column rank, i.e., for $\beta=0$, meaning that the $\boldsymbol{\theta}$ becomes identifiable when $h^{(1)}(z)$ and $h^{(2)}(z)$ share a common zero.

Fig. 3 shows the presented bounds in Fig. 2 with $\beta=0$. We see that Tr[AMVB$_{r,c}^{NCG}(\boldsymbol{\theta})$] can be larger than Tr[AMVB$_{r,c}^{CG}(\boldsymbol{\theta})$], depending on the SNR values. This interesting counterexample does not contradict neither Result 3 (due to the presence of nuisance parameters), nor Eq. (33) (due to the structure information of $\mathbf{R}_s$ and $\mathbf{C}_s$ that is taken into account).

Fig. 4 compares the AMVBs to the CRBs associated with BPSK and QPSK distribution sources. Because the associated PDF of $\mathbf{x}(n)$ is a mixture of c^{L+M+1} ($c=2$ [resp. 4] for BPSK [resp. QPSK] modulations), Gaussian PDFs:

$$p(\mathbf{x}(n);\boldsymbol{\theta},\boldsymbol{\alpha})=\frac{1}{c^{L+M+1}\pi^{P(L+1)}\sigma_e^{2P(L+1)}}\sum_{l=1}^{c^{L+M+1}}e^{-\|\mathbf{x}(n)-\mathbf{A}(\boldsymbol{\theta})\mathbf{s}_l\|^2/\sigma_e^2}$$

with $\mathbf{s}_l\overset{\text{def}}{=}\sigma_s e^{i\alpha_s}\boldsymbol{\epsilon}_l$

with $\boldsymbol{\epsilon}_l=(\epsilon_{1,l},\epsilon_{2,l},\ldots,\epsilon_{L+M+1,l})^T$, $l=1,\ldots,c^{L+M+1}$ where $\epsilon_{k,l}$ represent all the sequence of $L+M+1$ $\{-1,+1\}$ [resp. $\{-1,+1,-i,+i\}$] BPSK [resp. QPSK] symbols, this latter CRB appears to be prohibitive to compute. Thus we use a numerical approximation derived from the strong law of large numbers applied to the expectation of the first expressions of the different FIMs (14), (15), (19) and (20).

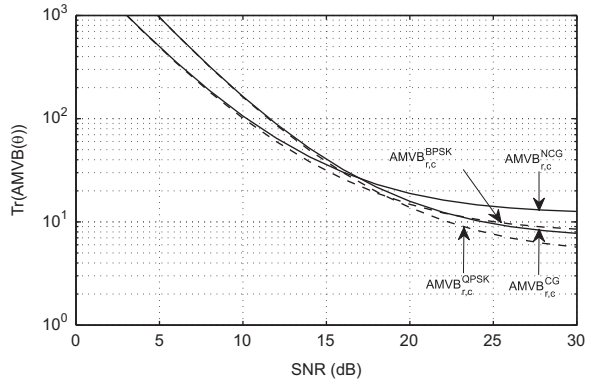

Fig. 3. Tr[AMVB$_{r,c}^{BPSK}(\boldsymbol{\theta})$], Tr[AMVB$_{r,c}^{QPSK}(\boldsymbol{\theta})$], Tr[AMVB$_{r,c}^{CG}(\boldsymbol{\theta})$] and Tr[AMVB$_{r,c}^{NCG}(\boldsymbol{\theta})$] as a function of the SNR for $\beta=0$.

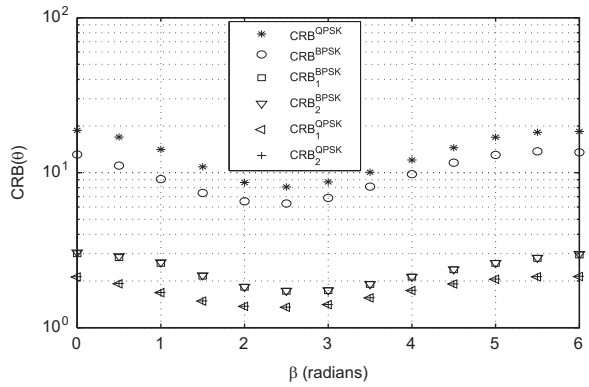

Fig. 4. Tr[CRB($\boldsymbol{\theta}$)], Tr[$(\mathbf{J}_\theta-\mathbf{J}_{*,\theta}\mathbf{J}_\theta^{-1}\mathbf{J}_{*,\theta}^*)^{-1}$] and Tr[$\mathbf{J}_\theta^{-1}$] for QPSK and BPSK modulations as a function of the phase parameter β for the channel.

Note that in contrast to Gaussian inputs, efficient algorithms are no longer circular distributed because here $\mathbf{J}_{*,\theta}\neq\mathbf{0}$. Fig. 4 exhibits the normalized ($N=1$) Tr[CRB$^{BPSK}(\boldsymbol{\theta})$] and Tr[CRB$^{QPSK}(\boldsymbol{\theta})$] with CRB($\boldsymbol{\theta}$) $=[(\mathbf{J}_{\tilde{\theta}}-\mathbf{J}_{\tilde{\theta},\alpha}\mathbf{J}_\alpha^{-1}\mathbf{J}_{\tilde{\theta},\alpha}^H)^{-1}]_{\theta,\theta}$ given by (18).It also exhibits Tr[CRB$_1^{BPSK}(\boldsymbol{\theta})$], Tr[CRB$_2^{BPSK}(\boldsymbol{\theta})$], Tr[CRB$_1^{QPSK}(\boldsymbol{\theta})$] and Tr[CRB$_2^{QPSK}(\boldsymbol{\theta})$], with CRB$_1(\boldsymbol{\theta})\overset{\text{def}}{=}(\mathbf{J}_\theta-\mathbf{J}_{*,\theta}\mathbf{J}_\theta^{-1}\mathbf{J}_{*,\theta}^*)^{-1}$ and CRB$_2(\boldsymbol{\theta})\overset{\text{def}}{=}\mathbf{J}_\theta^{-1}$ to see the impact of the nuisance parameters and $\mathbf{J}_{*,\theta}$ on the CRB. We see that Tr[CRB$^{QPSK}(\boldsymbol{\theta})$] is still larger than Tr[CRB$^{BPSK}(\boldsymbol{\theta})$]. The presence of unknown nuisance parameters degrades the trace

of the CRB of almost 10 dB, but the impact of the nonzero value of $\mathbf{J}_{*,\theta}$ has little influence of the CRBs.

This contrasts with the estimation of the gain matrix for complex ICA model [20], in which the nonzero value of $\mathbf{J}_{*,\theta}$ can have a strong impact on the CRB. This is shown in Figs. 5 and 6 that exhibit the normalized ($N=1$) Tr[CRB($\boldsymbol{\theta}$)] and Tr[CRB$_1$($\boldsymbol{\theta}$)] where here $\mathrm{CRB}(\boldsymbol{\theta}) \stackrel{\mathrm{def}}{=} (\mathbf{J}_\theta - \mathbf{J}_{*,\theta}\mathbf{J}_\theta^{-1}\mathbf{J}_{*,\theta}^*)^{-1}$ (no nuisance parameter) and $\mathrm{CRB}_1(\boldsymbol{\theta}) \stackrel{\mathrm{def}}{=} \mathbf{J}_\theta^{-1}$. In this experiment, $\mathbf{x}(n) = \mathbf{A}\mathbf{s}(n)$ where $\mathbf{A}$ is invertible square unstructured, $\boldsymbol{\theta} \stackrel{\mathrm{def}}{=} \mathrm{vec}(\mathbf{WA})$, where $\mathbf{W}$ is the estimated demixing matrix $\mathbf{A}^{-1}$ and $\mathbf{WA}$ is the so-called gain matrix. For these two figures, we consider 3 independent generalized Gaussian distributed sources with shape parameter $c > 0$ and noncircularity coefficient $\gamma \in [0, 1]$. We see in these figures that the nonzero terms $\mathbf{J}_{*,\theta}$ can have a large influence on the CRB, particularly for c close to 1 (Gaussian sources for which $\boldsymbol{\theta}$ is not identifiable) and for γ close to 0 (circular sources for which $\boldsymbol{\theta}$ is not identifiable) or 1 (rectilinear sources). This proves that the traditional lower bound $\mathbf{J}_\theta^{-1}$ on the CRB can be very loose.

7. Conclusion

Despite the real-valued nature of physical signals, complex-valued signals and parameters are generally encountered in many science and engineering problems as the complex formalism can provide a natural way to capture the physical characteristics of these signals and parameters. The wider deployment of complex-valued signal processing is still hindered by the fact that the statistical tools for handling complex-valued parameters are missing or scattered in the literature. This paper has provided a rigorous and unified framework to study the statistical performance of complex-valued parameter estimators, with a special attention to the complex Cramer–Rao and asymptotically minimum variance-type performance bounds where new extensions and properties have been presented with a special emphasis on noisy linear mixtures. Some of these results have been illustrated by numerical examples with blind identification of complex SIMO channels and complex independent component analysis examples and models.

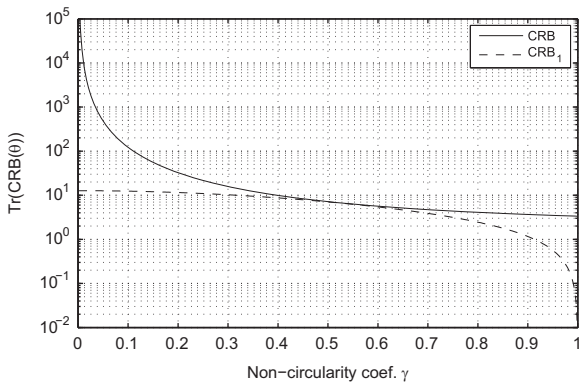

Fig. 5. Tr[CRB($\boldsymbol{\theta}$)] and Tr[$\mathbf{J}_\theta^{-1}$] as a function of the noncircularity coefficient γ for $c=0.25$.

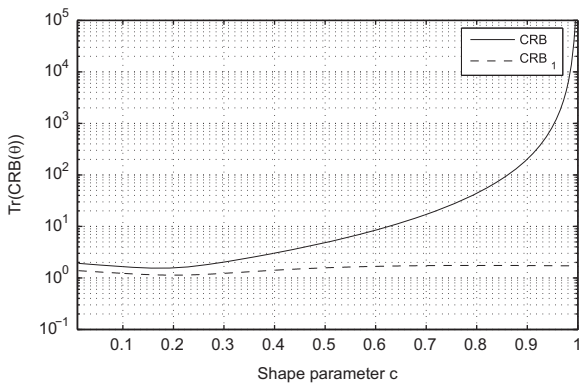

Fig. 6. Tr[CRB($\boldsymbol{\theta}$)] and Tr[$\mathbf{J}_\theta^{-1}$] as a function of the shape parameter c for $\gamma=0.9$.

Appendix

Proof of Eqs. (4) and (5). From (3) and (2), we get, respectively,

$$\tilde{\hat{\boldsymbol{\theta}}}_N - \tilde{\boldsymbol{\theta}} = \begin{pmatrix} \mathbf{D}_g & \mathbf{D}_{*g} \\ \mathbf{D}_g^* & \mathbf{D}_{*g}^* \end{pmatrix} (\tilde{\mathbf{s}}_N - \tilde{\mathbf{s}}) + o(\|\mathbf{s}_N - \mathbf{s})\|$$

and

$$\sqrt{N}\tilde{\mathbf{a}}^H(\tilde{\mathbf{s}}_N - \tilde{\mathbf{s}}) \xrightarrow{\mathcal{L}} \mathcal{N}_R\left(\mathbf{0}; \tilde{\mathbf{a}}^H \begin{pmatrix} \mathbf{R}_s & \mathbf{C}_s \\ \mathbf{C}_s^* & \mathbf{R}_s^* \end{pmatrix} \tilde{\mathbf{a}}\right),$$

for any $\mathbf{a} \in \mathbb{C}^p$. Then following the steps of the proof of the standard theorem of continuity [30, Th.B, p. 124], we deduce for any $\mathbf{b} \in \mathbb{C}^q$:

$$\sqrt{N}\tilde{\mathbf{b}}^H(\tilde{\hat{\boldsymbol{\theta}}}_N - \tilde{\boldsymbol{\theta}}) \xrightarrow{\mathcal{L}} \mathcal{N}_R\left(\mathbf{0}; \tilde{\mathbf{b}}^H \begin{pmatrix} \mathbf{D}_g & \mathbf{D}_{*g} \\ \mathbf{D}_g^* & \mathbf{D}_{*g}^* \end{pmatrix} \begin{pmatrix} \mathbf{R}_s & \mathbf{C}_s \\ \mathbf{C}_s^* & \mathbf{R}_s^* \end{pmatrix} \begin{pmatrix} \mathbf{D}_g & \mathbf{D}_{*g} \\ \mathbf{D}_g^* & \mathbf{D}_{*g}^* \end{pmatrix}^H \tilde{\mathbf{b}}\right). \quad \square$$

Proof of Eq. (9). If the mapping $\mathbf{g}$ is $\mathbb{R}$-differentiable to the second-order, the $\mathbb{CR}$-calculus [11] allows us to give the kth component of $\mathbf{g}$:

$$\begin{aligned}[\mathbf{g}(\mathbf{s}_N)]_k = {} & [\mathbf{g}(\mathbf{s})]_k + \frac{\partial \mathbf{g}_k}{\partial \mathbf{s}}(\mathbf{s}_N - \mathbf{s}) + \frac{\partial \mathbf{g}}{\partial \mathbf{s}^*}(\mathbf{s}_N - \mathbf{s})^* \\ & + \frac{1}{2}(\mathbf{s}_N - \mathbf{s})^H \frac{\partial}{\partial \mathbf{s}}\left(\frac{\partial \mathbf{g}_k}{\partial \mathbf{s}}\right)^H (\mathbf{s}_N - \mathbf{s}) + \frac{1}{2}(\mathbf{s}_N - \mathbf{s})^T \left[\frac{\partial}{\partial \mathbf{s}}\left(\frac{\partial \mathbf{g}_k}{\partial \mathbf{s}}\right)^H\right]^T (\mathbf{s}_N - \mathbf{s})^* \\ & + \frac{1}{2}(\mathbf{s}_N - \mathbf{s})^H \frac{\partial}{\partial \mathbf{s}^*}\left(\frac{\partial \mathbf{g}_k}{\partial \mathbf{s}}\right)^H (\mathbf{s}_N - \mathbf{s})^* + \frac{1}{2}(\mathbf{s}_N - \mathbf{s})^T \frac{\partial}{\partial \mathbf{s}}\left(\frac{\partial \mathbf{g}_k}{\partial \mathbf{s}}\right)^T (\mathbf{s}_N - \mathbf{s}) \\ & + o(\|\mathbf{s}_N - \mathbf{s}\|^2). \end{aligned} \tag{40}$$

Taking the expectation of (40) and assuming that the necessary mathematics conditions concerning the remainder are met (see comments in Section 2), it holds

$$\begin{aligned} \mathrm{E}[\hat{\boldsymbol{\theta}}_N]_k = {} & \theta_k \\ & + \mathrm{Tr}\left(\mathrm{E}\left((\mathbf{s}_N - \mathbf{s})(\mathbf{s}_N - \mathbf{s})^H\right)\left(\frac{\partial}{\partial \mathbf{s}}\left(\frac{\partial \mathbf{g}_k}{\partial \mathbf{s}}\right)^H\right)\right) \\ & + \frac{1}{2}\mathrm{Tr}\left(\mathrm{E}\left((\mathbf{s}_N - \mathbf{s})^*(\mathbf{s}_N - \mathbf{s})^H\right)\left(\frac{\partial}{\partial \mathbf{s}^*}\left(\frac{\partial \mathbf{g}_k}{\partial \mathbf{s}}\right)^H\right)\right) \\ & + \frac{1}{2}\mathrm{Tr}\left(\mathrm{E}\left((\mathbf{s}_N - \mathbf{s})(\mathbf{s}_N - \mathbf{s})^T\right)\left(\frac{\partial}{\partial \mathbf{s}}\left(\frac{\partial \mathbf{g}_k}{\partial \mathbf{s}}\right)^T\right)\right) + o\left(\frac{1}{N}\right). \end{aligned}$$

Eq. (8) concludes the proof. $\square$

Proof of Result 1. From the $\mathbb{R}$-differentiability of the function $\mathbf{g}$, we get from (3) the augmented equality:

$$\tilde{\mathbf{g}}(\mathbf{s}+\delta\mathbf{s}) = \tilde{\boldsymbol{\theta}} + \tilde{\mathbf{D}}_g \delta\tilde{\mathbf{s}} + o(\|\delta\mathbf{s}\|). \tag{41}$$

In addition, because $\tilde{\mathbf{g}}[\mathbf{s}(\boldsymbol{\theta})] = \tilde{\boldsymbol{\theta}}$ for all $\boldsymbol{\theta}$, we have

$$\begin{aligned}\tilde{\mathbf{g}}[\mathbf{s}(\boldsymbol{\theta}+\delta\boldsymbol{\theta})] &= \tilde{\boldsymbol{\theta}} + \delta\tilde{\boldsymbol{\theta}} \\ &= \tilde{\mathbf{g}}[\mathbf{s}(\boldsymbol{\theta}) + (\mathbf{D}_s, \mathbf{D}_{*,s})\delta\tilde{\boldsymbol{\theta}} + o(\|\delta\boldsymbol{\theta}\|)] \\ &= \tilde{\boldsymbol{\theta}} + \tilde{\mathbf{D}}_g \tilde{\mathbf{D}}_s \delta\tilde{\boldsymbol{\theta}} + o(\|\delta\boldsymbol{\theta}\|),\end{aligned}$$

where we have used the $\mathbb{R}$-differentiability of the functions $\boldsymbol{\theta} \longmapsto \mathbf{s}(\boldsymbol{\theta})$ and $\mathbf{s} \longmapsto \mathbf{g}(\mathbf{s})$ in the second and third equalities, respectively. Therefore $\tilde{\mathbf{D}}_g$ is a left inverse of $\tilde{\mathbf{D}}_s$, i.e., $\tilde{\mathbf{D}}_g \tilde{\mathbf{D}}_s = \mathbf{I}_{2q}$. So it is easy to check that this implies the following equality:

$$\begin{aligned}&\tilde{\mathbf{D}}_g \mathbf{R}_{\tilde{s}} \tilde{\mathbf{D}}_g^H - (\tilde{\mathbf{D}}_s^H \mathbf{R}_{\tilde{s}}^{-1} \tilde{\mathbf{D}}_s)^{-1} \\ &= [\tilde{\mathbf{D}}_g - (\tilde{\mathbf{D}}_s^H \mathbf{R}_{\tilde{s}}^{-1} \tilde{\mathbf{D}}_s)^{-1} \tilde{\mathbf{D}}_s^H \mathbf{R}_{\tilde{s}}^{-1}] \mathbf{R}_{\tilde{s}} [\tilde{\mathbf{D}}_g - (\tilde{\mathbf{D}}_s^H \mathbf{R}_{\tilde{s}}^{-1} \tilde{\mathbf{D}}_s)^{-1} \tilde{\mathbf{D}}_s^H \mathbf{R}_{\tilde{s}}^{-1}]^H,\end{aligned}$$

that concludes the proof of (10). □

If $V_N(\boldsymbol{\beta}) \stackrel{\text{def}}{=} [\tilde{\mathbf{s}}_N - \tilde{\mathbf{s}}(\boldsymbol{\beta})]^H \mathbf{W}_N [\tilde{\mathbf{s}}_N - \tilde{\mathbf{s}}(\boldsymbol{\beta})]$, its $\mathbb{R}$-derivative $\partial V_N(\boldsymbol{\beta})/\partial\boldsymbol{\beta}$ is zero at $\boldsymbol{\beta} = \boldsymbol{\theta} + \delta\boldsymbol{\theta}$ where $\boldsymbol{\theta} + \delta\boldsymbol{\theta}$ is associated with $\tilde{\mathbf{s}}_N = \tilde{\mathbf{s}} + \delta\tilde{\mathbf{s}}$. Expanding this derivative by a perturbation analysis and using $\tilde{\mathbf{s}}_N - \tilde{\mathbf{s}}(\boldsymbol{\beta}) = \delta\tilde{\mathbf{s}} - \tilde{\mathbf{D}}_s \delta\tilde{\boldsymbol{\theta}} + o(\|\delta\boldsymbol{\theta}\|)$, we straightforwardly obtain $(\tilde{\mathbf{D}}_s^H \mathbf{R}_{\tilde{s}}^{-1} \tilde{\mathbf{D}}_s)\delta\tilde{\boldsymbol{\theta}} + o(\|\delta\boldsymbol{\theta}\|) = \tilde{\mathbf{D}}_s^H \mathbf{R}_{\tilde{s}}^{-1} \delta\tilde{\mathbf{s}} + o(\|\delta\|\mathbf{s}\|)$. Consequently, the algorithm $\mathbf{g}$ defined by (11) satisfies

$$\tilde{\mathbf{g}}(\tilde{\mathbf{s}} + \delta\tilde{\mathbf{s}}) = \tilde{\boldsymbol{\theta}} + (\tilde{\mathbf{D}}_s^H \mathbf{R}_{\tilde{s}}^{-1} \tilde{\mathbf{D}}_s)^{-1} \tilde{\mathbf{D}}_s^H \mathbf{R}_{\tilde{s}}^{-1} \delta\tilde{\mathbf{s}} + o(\|\delta\mathbf{s})\|).$$

Consequently, the $\mathbb{C}$-derivative of the mapping $\tilde{\mathbf{s}} \longmapsto \tilde{\boldsymbol{\theta}} = \tilde{\mathbf{g}}(\tilde{\mathbf{s}})$ involved by (11) is $\tilde{\mathbf{D}}_g = (\tilde{\mathbf{D}}_s^H \mathbf{R}_{\tilde{s}}^{-1} \tilde{\mathbf{D}}_s)^{-1} \tilde{\mathbf{D}}_s^H \mathbf{R}_{\tilde{s}}^{-1}$ and the covariance of the asymptotic distribution of $\tilde{\boldsymbol{\theta}}_N$ is therefore $\mathbf{R}_{\tilde{\theta}} = \tilde{\mathbf{D}}_g \mathbf{R}_{\tilde{s}} \tilde{\mathbf{D}}_g^H = (\tilde{\mathbf{D}}_s^H \mathbf{R}_{\tilde{s}}^{-1} \tilde{\mathbf{D}}_s)^{-1}$ that concludes the proof of (11). □

Proof of Result 2. Whittle formula (24) applies to $\overline{\mathbf{x}}(n)$ associated with the real-valued parameter $\overline{\boldsymbol{\theta}}$ where $\mathbf{U} \stackrel{\text{def}}{=} \frac{1}{2}\begin{bmatrix}\mathbf{I} & \mathbf{I} \\ -i\mathbf{I} & i\mathbf{I}\end{bmatrix}$ of conformable dimension.

Using

$$\mathbf{D}_{\mathbf{r}_{\tilde{x}}}(f) = \mathbf{U}_r \frac{\partial \mathbf{r}_{\tilde{x}}(f)}{\partial \overline{\boldsymbol{\theta}}} = \mathbf{U}_r \left[\frac{\partial \mathbf{r}_{\tilde{x}}(f)}{\partial \boldsymbol{\theta}}, \frac{\partial \mathbf{r}_{\tilde{x}}(f)}{\partial \boldsymbol{\theta}^*}\right] \mathbf{U}_q^{-1} = \mathbf{U}_r [\mathbf{D}_{\mathbf{r}_{\tilde{x}}}(f), \mathbf{D}_{*,\mathbf{r}_{\tilde{x}}}(f)] \mathbf{U}_q^{-1}$$

and $\mathbf{R}_{\overline{x}}(f) = \mathbf{U}_r \mathbf{R}_{\tilde{x}}(f) \mathbf{U}_r^H$, we get from $\mathbf{R}_{\hat{\overline{\theta}}} \geq \mathbf{J}_{\overline{\theta}}^{-1}$ after straightforward algebra manipulations:

$$\begin{aligned}\mathbf{U}_q \mathbf{R}_{\hat{\overline{\theta}}} \mathbf{U}_q^H \geq \mathbf{U}_q &\left(\frac{N}{2}\int_{-1/2}^{+1/2} \begin{bmatrix}\mathbf{D}_{\mathbf{r}_{\tilde{x}}}^H(f) \\ \mathbf{D}_{*,\mathbf{r}_{\tilde{x}}}^H(f)\end{bmatrix} \left(\mathbf{R}_x^{-*}(f) \otimes \mathbf{R}_x^{-1}(f)\right)\right. \\ &\left.\times [\mathbf{D}_{\mathbf{r}_{\tilde{x}}}(f), \mathbf{D}_{*,\mathbf{r}_{\tilde{x}}}(f)] df,\right)^{-1} \mathbf{U}_q^H,\end{aligned}$$

that concludes the proof of (25). □

Proof of Result 3. From (29),

$$\mathbf{J}_{\tilde{\theta}}(\mathbf{m}_x, \mathbf{R}_x, \mathbf{C}_x) = \left[\tilde{\mathbf{D}}_{s_1}^H, \tilde{\mathbf{D}}_{s_2}^H\right] \mathbf{R}_{\tilde{s}}^{-1} \begin{bmatrix}\tilde{\mathbf{D}}_{s_1} \\ \tilde{\mathbf{D}}_{s_2}\end{bmatrix},$$

where $\tilde{\mathbf{s}}$ is split in $\tilde{\mathbf{s}}_1$ and $\tilde{\mathbf{s}}_2$, i.e., $\tilde{\mathbf{s}} = \begin{bmatrix}\tilde{\mathbf{s}}_1 \\ \tilde{\mathbf{s}}_2\end{bmatrix}$ with $\tilde{\mathbf{s}}_1 = [\mathbf{m}_x^T, \mathbf{m}_x^H, \text{vec}^T(\mathbf{R}_x)]^T$ and $\tilde{\mathbf{s}}_2 = [\text{v}^T(\mathbf{C}_x), \text{v}^H(\mathbf{C}_x)]^T$, and where $\tilde{\mathbf{D}}_{s_i} \stackrel{\text{def}}{=} \begin{bmatrix}\mathbf{D}_{\tilde{s}_i} & \mathbf{D}_{*,\tilde{s}_i} \\ \mathbf{D}_{*,\tilde{s}_i}^* & \mathbf{D}_{\tilde{s}_i}^*\end{bmatrix}$, $i = 1,2$.

Consequently $\mathbf{J}_{\tilde{\theta}}(\mathbf{m}_x, \mathbf{R}_x, \mathbf{0}) = \tilde{\mathbf{D}}_{s_1}^H \mathbf{R}_{\tilde{s}_1}^{-1} \tilde{\mathbf{D}}_{s_1}$ where $\mathbf{R}_{\tilde{s}} = \begin{bmatrix}\mathbf{R}_{\tilde{s}_1} & \mathbf{R}_{\tilde{s}_{1,2}} \\ \mathbf{R}_{\tilde{s}_{1,2}}^H & \mathbf{R}_{\tilde{s}_2}\end{bmatrix}$. From lemma [31, A.4], $\left[\tilde{\mathbf{D}}_{s_1}^H, \tilde{\mathbf{D}}_{s_2}^H\right] \mathbf{R}_{\tilde{s}}^{-1} \begin{bmatrix}\tilde{\mathbf{D}}_{s_1} \\ \tilde{\mathbf{D}}_{s_2}\end{bmatrix} \geq \tilde{\mathbf{D}}_{s_1}^H \mathbf{R}_{\tilde{s}_1}^{-1} \tilde{\mathbf{D}}_{s_1}$, that concludes the proof of (30). □

Proof of Eq. (32). In the circular case, all the steps of the proofs given for the DOA model in [24] remain valid with the general model (31), where [24, rel. (16)] is replaced by

$$\frac{\partial \mathbf{R}_x}{\partial \theta_k} = \frac{\partial \mathbf{A}(\boldsymbol{\theta})}{\partial \theta_k} \mathbf{R}_s \mathbf{A}^H(\boldsymbol{\theta}) + \mathbf{A}(\boldsymbol{\theta}) \mathbf{R}_s \frac{\partial \mathbf{A}^H(\boldsymbol{\theta})}{\partial \theta_k}$$

and where the term $\mathbf{A}\mathbf{c}_k \mathbf{d}_k^H$ in [24, rel. (18)] and [24, rel. (27)] is replaced by the term $\mathbf{A}(\boldsymbol{\theta}) \mathbf{R}_s \partial \mathbf{A}^H(\boldsymbol{\theta})/\partial\theta_k$.

In the noncircular case, the proof follows the steps of [21, Appendix C] based on [24], where [24, rel. (16)] is replaced by

$$\frac{\partial \mathbf{R}_{\tilde{x}}}{\partial \theta_k} = \frac{\partial \tilde{\mathbf{A}}(\boldsymbol{\theta})}{\partial \theta_k} \mathbf{R}_{\tilde{s}} \tilde{\mathbf{A}}^H(\boldsymbol{\theta}) + \tilde{\mathbf{A}}(\boldsymbol{\theta}) \mathbf{R}_{\tilde{s}} \frac{\partial \tilde{\mathbf{A}}^H(\boldsymbol{\theta})}{\partial \theta_k} \quad \text{with } \tilde{\mathbf{A}}(\boldsymbol{\theta}) \stackrel{\text{def}}{=} \begin{bmatrix}\mathbf{A}(\boldsymbol{\theta}) & \mathbf{0} \\ \mathbf{0} & \mathbf{A}^*(\boldsymbol{\theta})\end{bmatrix}$$

and where the term $\mathbf{A}\mathbf{c}_k \mathbf{d}_k^H$ in [24, rel. (18)] and [24, rel. (27)] is replaced by the term $\tilde{\mathbf{A}}(\boldsymbol{\theta}) \mathbf{R}_{\tilde{s}} \partial \tilde{\mathbf{A}}^H(\boldsymbol{\theta})/\partial\theta_k$. □

Proof of Result 4. Using $\overline{\boldsymbol{\theta}} = \mathbf{U}\tilde{\boldsymbol{\theta}}$, we have $\text{CRB}(\tilde{\boldsymbol{\theta}}) = [\mathbf{U}^H \text{CRB}^{-1}(\overline{\boldsymbol{\theta}}) \mathbf{U}]^{-1}$. Replacing $\text{CRB}^{-1}(\overline{\boldsymbol{\theta}})$ from (32) by $8N/\sigma_e^2 [\text{Re}\left(\mathbf{U}(\partial\mathbf{a}/\partial\boldsymbol{\theta}, \partial\mathbf{a}/\partial\boldsymbol{\theta}^*)^H \left(\mathbf{H}^T \otimes \boldsymbol{\Pi}_A^\perp\right)(\partial\mathbf{a}/\partial\boldsymbol{\theta}, \partial\mathbf{a}/\partial\boldsymbol{\theta}^*) \mathbf{U}^H\right)]$, (33)–(35) are obtained from straightforward manipulations. □

Proof of Eq. (37). First, from [31, Lemma A.4], we have $\mathbf{H}_{\text{nc}} - \mathbf{H}_{\text{c}} \geq \mathbf{0}$ with $\mathbf{H}_{\text{nc}} \stackrel{\text{def}}{=} \left[\mathbf{R}_s \mathbf{A}^H, \mathbf{C}_s \mathbf{A}^T\right] \mathbf{R}_{\tilde{x}}^{-1} \begin{bmatrix}\mathbf{A}\mathbf{R}_s \\ \mathbf{A}^* \mathbf{C}_s^*\end{bmatrix}$ and $\mathbf{H}_{\text{c}} \stackrel{\text{def}}{=} \mathbf{R}_s \mathbf{A}^H \mathbf{R}_x^{-1} \mathbf{A} \mathbf{R}_s$, and this inequality applies to the transpose of these matrices: $\mathbf{H}_{\text{nc}}^T - \mathbf{H}_{\text{c}}^T \geq \mathbf{0}$. Then, because $\boldsymbol{\Pi}_A^\perp \geq \mathbf{0}$, $(\mathbf{H}_{\text{nc}}^T - \mathbf{H}_{\text{c}}^T) \otimes \boldsymbol{\Pi}_A^\perp \geq \mathbf{0}$ thanks to a standard result of linear algebra (see, e.g., [32, prop. 11.5]. Consequently $\partial\mathbf{a}^H/\partial\overline{\boldsymbol{\theta}}\left((\mathbf{H}_{\text{nc}}^T - \mathbf{H}_{\text{c}}^T) \otimes \boldsymbol{\Pi}_A^\perp)\right)\partial\mathbf{a}/\partial\overline{\boldsymbol{\theta}} \geq \mathbf{0}$. This inequality is extended to its conjugate and consequently to the real-valued symmetric matrix $\text{Re}\left(\partial\mathbf{a}^H/\partial\overline{\boldsymbol{\theta}}\left((\mathbf{H}_{\text{nc}}^T - \mathbf{H}_{\text{c}}^T) \otimes \boldsymbol{\Pi}_A^\perp)\right)\partial\mathbf{a}/\partial\overline{\boldsymbol{\theta}}\right)$. Then by inversion

$$\begin{aligned}&\left[\text{Re}\left(\partial\mathbf{a}^H/\partial\overline{\boldsymbol{\theta}}\left(\mathbf{H}_{\text{nc}}^T \otimes \boldsymbol{\Pi}_A^\perp\right)\partial\mathbf{a}/\partial\overline{\boldsymbol{\theta}}\right)\right]^{-1} \\ &\quad \leq \left[\text{Re}\left(\partial\mathbf{a}^H/\partial\overline{\boldsymbol{\theta}}\left(\mathbf{H}_{\text{c}}^T \otimes \boldsymbol{\Pi}_A^\perp\right)\partial\mathbf{a}/\partial\overline{\boldsymbol{\theta}}\right)\right]^{-1}\end{aligned}$$

and consequently $\text{CRB}_{\mathbf{R}_x, \mathbf{C}_x}(\overline{\boldsymbol{\theta}}) \leq \text{CRB}_{\mathbf{R}_x, \mathbf{0}}(\overline{\boldsymbol{\theta}})$.

Using the one-to-one linear mapping $\overline{\boldsymbol{\theta}} = \mathbf{U}\tilde{\boldsymbol{\theta}}$ where $\mathbf{U}$ is unitary ($\mathbf{U}^{-1} = 2\mathbf{U}^H$), $\text{CRB}(\tilde{\boldsymbol{\theta}}) = 4\mathbf{U}^H \text{CRB}(\overline{\boldsymbol{\theta}}) \mathbf{U}$. Consequently $\text{CRB}_{\mathbf{R}_x, \mathbf{C}_x}(\tilde{\boldsymbol{\theta}}) \leq \text{CRB}_{\mathbf{R}_x, \mathbf{0}}(\tilde{\boldsymbol{\theta}})$, as well. □

Proof of Eq. (38). First, note that the AMVB derivations in [27] apply to the real-valued parameter $\overline{\boldsymbol{\theta}}$ and thus we have

$$\mathbf{R}_{\overline{\theta}} \geq \text{AMVB}_s(\overline{\theta}) \stackrel{\text{def}}{=} (\mathbf{D}_{s,\overline{\theta}}^H \mathbf{R}_s^{\#} \mathbf{D}_{s,\overline{\theta}})^{-1} \tag{42}$$

where $\mathbf{D}_{s,\overline{\theta}} \stackrel{\text{def}}{=} d\mathbf{s}(\overline{\boldsymbol{\theta}})/d\overline{\boldsymbol{\theta}}$ which is related to the $\mathbb{R}$-derivative $\mathbf{D}_s = \partial\mathbf{s}/\partial\boldsymbol{\theta}$ and the conjugate $\mathbb{R}$-derivative $\mathbf{D}_{*,s} = \partial\mathbf{s}/\partial\boldsymbol{\theta}^*$ of $\mathbf{s}$ by $\mathbf{D}_{s,\overline{\theta}} = [\mathbf{D}_s, \mathbf{D}_{*,s}]\mathbf{U}^{-1}$. Using $\overline{\boldsymbol{\theta}} = \mathbf{U}\tilde{\boldsymbol{\theta}}$, $\mathbf{R}_{\overline{\theta}} = \mathbf{U}\mathbf{R}_{\tilde{\theta}}\mathbf{U}^H$, (42) is equivalent to

$$\mathbf{R}_{\tilde{\theta}} \geq [(\mathbf{D}_s, \mathbf{D}_{*,s})^H \mathbf{R}_s^{\#} (\mathbf{D}_s, \mathbf{D}_{*,s})]^{-1}.$$

Consider now the statistic $\mathbf{s}_N = \text{vec}(\boldsymbol{\Pi}_{R_x,N})$. Using the Hermitian structure of $\boldsymbol{\Pi}_{R_x,N}$, its asymptotic covariance $\mathbf{R}_s$ and complementary covariance $\mathbf{C}_s$ are related by

$$\mathbf{R}_{\tilde{s}} = \begin{bmatrix} \mathbf{R}_s & \mathbf{C}_s \\ \mathbf{C}_s^* & \mathbf{R}_s^* \end{bmatrix} = 2\mathbf{B}_K \mathbf{R}_s \mathbf{B}_K^H,$$

where $\mathbf{B}_K \stackrel{\text{def}}{=} \frac{1}{\sqrt{2}}\begin{bmatrix} \mathbf{I} \\ \mathbf{K} \end{bmatrix}$ satisfying $\mathbf{B}_K^H \mathbf{B}_K = \mathbf{I}$. Consequently $\mathbf{R}_{\tilde{s}}^{\#} = \frac{1}{2}\mathbf{B}_K \mathbf{R}_s^{\#} \mathbf{B}_K^H$ from [32, Prop. 7.69]. This implies

$$\begin{aligned} \tilde{\mathbf{D}}_s^H \mathbf{R}_{\tilde{s}}^{\#} \tilde{\mathbf{D}}_s &= \frac{1}{4}\begin{bmatrix} \mathbf{D}_s^H & \mathbf{D}_{*,s}^T \\ \mathbf{D}_{*,s}^H & \mathbf{D}_s^T \end{bmatrix} \begin{bmatrix} \mathbf{I} \\ \mathbf{K} \end{bmatrix} \mathbf{R}_s^{\#} [\mathbf{I}, \mathbf{K}] \begin{bmatrix} \mathbf{D}_s & \mathbf{D}_{*,s} \\ \mathbf{D}_{*,s}^* & \mathbf{D}_s^* \end{bmatrix} \\ &= (\mathbf{D}_s, \mathbf{D}_{*,s})^H \mathbf{R}_s^{\#} (\mathbf{D}_s, \mathbf{D}_{*,s}), \end{aligned} \tag{43}$$

where $\mathbf{K}\mathbf{D}_s = \mathbf{D}_{*,s}^*$ and $\mathbf{K}\mathbf{D}_{*,s} = \mathbf{D}_s^*$ (due to the Hermitian structure of $\mathbf{s}_N = \boldsymbol{\Pi}_{R_x,N}$ and the relation $(d\mathbf{s}/d\boldsymbol{\theta})^* = d\mathbf{s}^*/d\boldsymbol{\theta}^*$ between $\mathbb{R}$-derivatives) are used in the second equality. So (38) is proved for $\mathbf{s}_N = \text{vec}(\boldsymbol{\Pi}_{R_x,N})$. The proofs for $\text{vec}(\boldsymbol{\Pi}_{R_{\tilde{x}},N})$ and $\text{vec}(\boldsymbol{\Pi}_{R_x,N}, \boldsymbol{\Pi}_{C_x,N})$ are similar. Finally, note that AMVB derivations in [27], the AMVB based on $\boldsymbol{\Pi}_{R_{\tilde{x}},N}$ and $(\boldsymbol{\Pi}_{R_x,N}, \boldsymbol{\Pi}_{C_x,N})$ coincide for $\overline{\boldsymbol{\theta}}$, and thus for $\tilde{\boldsymbol{\theta}}$. □

Proof of Result 5. Consider the statistic $\mathbf{s}_N = \text{vec}(\boldsymbol{\Pi}_{R_x,N})$ whose Moore–Penrose inverse of the covariance of its asymptotic distribution is given from [27] by

$$\mathbf{R}_s^{\#} = \frac{1}{\sigma_e^2}\left[(\mathbf{A}^*\mathbf{H}^*\mathbf{A}^T \otimes \boldsymbol{\Pi}_A^{\perp}) + (\boldsymbol{\Pi}_A^{\perp^T} \otimes \mathbf{A}\mathbf{H}\mathbf{A}^H)\right].$$

So from (43), $\tilde{\mathbf{D}}_s^H \mathbf{R}_{\tilde{s}}^{\#} \tilde{\mathbf{D}}_s$ is given by

$$\tilde{\mathbf{D}}_s^H \mathbf{R}_{\tilde{s}}^{\#} \tilde{\mathbf{D}}_s = \frac{1}{\sigma_e^2}\begin{bmatrix} \mathbf{D}_s^H \\ \mathbf{D}_{*,s}^H \end{bmatrix} \left[(\mathbf{A}^*\mathbf{H}^*\mathbf{A}^T \otimes \boldsymbol{\Pi}_A^{\perp}) + (\boldsymbol{\Pi}_A^{\perp^T} \otimes \mathbf{A}\mathbf{H}\mathbf{A}^H)\right] [\mathbf{D}_s, \mathbf{D}_{*,s}], \tag{44}$$

whose term (k,l) of the 1,1 block is written as

$$\frac{1}{\sigma_e^2}\text{vec}^H\left(\frac{\partial \boldsymbol{\Pi}_A^{\perp}}{\partial \theta_k}\right)\left((\mathbf{A}^*\mathbf{H}^*\mathbf{A}^T \otimes \boldsymbol{\Pi}_A^{\perp}) + (\boldsymbol{\Pi}_A^{\perp *} \otimes \mathbf{A}\mathbf{H}\mathbf{A}^H)\right)\text{vec}\left(\frac{\partial \boldsymbol{\Pi}_A^{\perp}}{\partial \theta_l}\right). \tag{45}$$

Using $\text{vec}^H(\partial\boldsymbol{\Pi}_A^{\perp}/\partial\theta_k) = \text{vec}^T\left((\partial\boldsymbol{\Pi}_A^{\perp}/\partial\theta_k^*)^T\right)$ and the identity $\text{Tr}(\mathbf{ABCD}) = \text{vec}^T(\mathbf{A}^T)(\mathbf{D}^T \otimes \mathbf{B})\text{vec}(\mathbf{C})$, the term (45) becomes

$$\begin{aligned} &\frac{1}{\sigma_e^2}\text{Tr}\left(\frac{\partial \boldsymbol{\Pi}_A^{\perp}}{\partial \theta_k^*}\boldsymbol{\Pi}_A^{\perp}\frac{\partial \boldsymbol{\Pi}_A^{\perp}}{\partial \theta_l}\mathbf{A}\mathbf{H}\mathbf{A}^H + \frac{\partial \boldsymbol{\Pi}_A^{\perp}}{\partial \theta_k^*}\mathbf{A}\mathbf{H}\mathbf{A}^H\frac{\partial \boldsymbol{\Pi}_A^{\perp}}{\partial \theta_l}\boldsymbol{\Pi}_A^{\perp}\right) \\ &= \frac{1}{\sigma_e^2}\text{Tr}\left(\underbrace{\mathbf{A}^H\frac{\partial \boldsymbol{\Pi}_A^{\perp}}{\partial \theta_k^*}\boldsymbol{\Pi}_A^{\perp}}\underbrace{\frac{\partial \boldsymbol{\Pi}_A^{\perp}}{\partial \theta_l}\mathbf{A}\mathbf{H}} + \underbrace{\frac{\partial \boldsymbol{\Pi}_A^{\perp}}{\partial \theta_k^*}\mathbf{A}\mathbf{H}\mathbf{A}^H}\underbrace{\frac{\partial \boldsymbol{\Pi}_A^{\perp}}{\partial \theta_l}\boldsymbol{\Pi}_A^{\perp}}\right). \end{aligned} \tag{46}$$

Then $\boldsymbol{\Pi}_A^{\perp}\mathbf{A} = \mathbf{0}$ implying $(\partial\boldsymbol{\Pi}_A^{\perp}/\partial\theta_i)\mathbf{A} + \boldsymbol{\Pi}_A^{\perp}\partial\mathbf{A}/\partial\theta_i = \mathbf{0}$ and $(\partial\boldsymbol{\Pi}_A^{\perp}/\partial\theta_i^*)\mathbf{A} + \boldsymbol{\Pi}_A^{\perp}\partial\mathbf{A}/\partial\theta_i^* = \mathbf{0}$, $i = k, l$ and using the relation $(\partial \times /\partial\theta_i)^* = \partial\times^*/\partial\theta_i^*$, $i = k, l$ between $\mathbb{R}$-derivatives of $\mathbf{A}$ and $\boldsymbol{\Pi}_A^{\perp}$, the term (46) is written as

$$\begin{aligned} &\frac{1}{\sigma_e^2}\text{Tr}\left(\left(\frac{\partial \mathbf{A}}{\partial \theta_k}\right)^H \boldsymbol{\Pi}_A^{\perp}\frac{\partial \mathbf{A}}{\partial \theta_l}\mathbf{H} + \left(\frac{\partial \mathbf{A}}{\partial \theta_k^*}\right)^T \boldsymbol{\Pi}_A^{\perp^T}\left(\frac{\partial \mathbf{A}}{\partial \theta_l^*}\right)^* \mathbf{H}^T\right) \\ &= \frac{1}{\sigma_e^2}\text{vec}^T\left(\frac{\partial \mathbf{A}}{\partial \theta_k}\right)^*\left(\mathbf{H}^T \otimes \boldsymbol{\Pi}_A^{\perp}\right)\text{vec}\left(\frac{\partial \mathbf{A}}{\partial \theta_l}\right) \\ &+ \frac{1}{\sigma_e^2}\text{vec}^T\left(\frac{\partial \mathbf{A}}{\partial \theta_k^*}\right)\left(\mathbf{H} \otimes \boldsymbol{\Pi}_A^{\perp^T}\right)\text{vec}\left(\frac{\partial \mathbf{A}}{\partial \theta_l^*}\right)^*. \end{aligned}$$

Consequently the block (1,1) of $\tilde{\mathbf{D}}_s^H \mathbf{R}_{\tilde{s}}^{\#} \tilde{\mathbf{D}}_s$ is given by

$$\frac{1}{\sigma_e^2}\left(\frac{\partial \mathbf{a}}{\partial \boldsymbol{\theta}}\right)^H\left(\mathbf{H}^T \otimes \boldsymbol{\Pi}_A^{\perp}\right)\left(\frac{\partial \mathbf{a}}{\partial \boldsymbol{\theta}}\right) + \frac{1}{\sigma_e^2}\left(\frac{\partial \mathbf{a}}{\partial \boldsymbol{\theta}^*}\right)^T\left(\mathbf{H} \otimes \boldsymbol{\Pi}_A^{\perp^T}\right)\left(\frac{\partial \mathbf{a}}{\partial \boldsymbol{\theta}^*}\right)^*,$$

which is equal to the block $\mathbf{J}_\theta$ of (34) for $N=1$. The derivation of the other three blocks of $\tilde{\mathbf{D}}_s^H \mathbf{R}_{\tilde{s}}^{\#} \tilde{\mathbf{D}}_s$ is obtained following the same steps and (39) is proved for the statistic $\mathbf{s}_N = \text{vec}(\boldsymbol{\Pi}_{R_x,N})$.

Concerning the statistics $\mathbf{s}_N = \text{vec}(\boldsymbol{\Pi}_{R_{\tilde{x}},N})$ and $\mathbf{s}_N = \text{vec}(\boldsymbol{\Pi}_{R_x,N}, \boldsymbol{\Pi}_{C_x,N})$, the covariance $\mathbf{R}_s$ of their asymptotic distribution and the associated Moore–Penrose inverse $\mathbf{R}_s^{\#}$ have been derived in [27]. Following the same steps that for $\mathbf{s}_N = \text{vec}(\boldsymbol{\Pi}_{R_x,N})$, (39) is proved for these other two statistics. □

Proof of Remark 1. Using an arbitrary square root $\mathbf{L}$ of $\boldsymbol{\Sigma}$, i.e., $\boldsymbol{\Sigma} = \mathbf{L}\mathbf{L}^H$, the model (31) becomes

$$\mathbf{x}_L(n) \stackrel{\text{def}}{=} \mathbf{L}^{-1}\mathbf{x}(n) = \mathbf{A}_L(\boldsymbol{\theta})\mathbf{s}(n) + \mathbf{e}_L(n) \tag{47}$$

with $\mathbf{A}_L(\boldsymbol{\theta}) \stackrel{\text{def}}{=} \mathbf{L}^{-1}\mathbf{A}(\boldsymbol{\theta})$ and $\mathbf{e}_L(n) \stackrel{\text{def}}{=} \mathbf{L}^{-1}\mathbf{e}(n)$. Consequently, the three conditions introduced in the beginning of Section 5 are still valid, and thus also all the results of this section apply by replacing $\mathbf{A}(\boldsymbol{\theta})$ by $\mathbf{L}^{-1}\mathbf{A}(\boldsymbol{\theta})$ in expressions (32) and (34)–(36).

Note that in these expressions, $\mathbf{a}$ and $\boldsymbol{\Pi}_A^{\perp}$ become $\mathbf{a}_L = \text{vec}(\mathbf{L}^{-1}\mathbf{A}) = \text{vec}(\mathbf{L}^{-1}\mathbf{A}\mathbf{I}) = (\mathbf{I} \otimes \mathbf{L}^{-1})\text{vec}(\mathbf{A}) = (\mathbf{I} \otimes \mathbf{L}^{-1})\mathbf{a}$ and $\boldsymbol{\Pi}_{A_L}^{\perp} = \mathbf{I} - \mathbf{L}^{-1}\mathbf{A}(\mathbf{A}^H\boldsymbol{\Sigma}^{-1}\mathbf{A})^{-1}\mathbf{A}^H\mathbf{L}^{-H}$, respectively. $\mathbf{H}$ is invariant in the circular and noncircular cases as

$$\mathbf{R}_s\mathbf{A}^H\mathbf{L}^{-H}(\mathbf{L}^{-1}\mathbf{R}_x\mathbf{L}^{-H})^{-1}\mathbf{L}^{-1}\mathbf{A}\mathbf{R}_s = \mathbf{H}$$

and

$$\begin{aligned} &[\mathbf{R}_s\mathbf{A}^H\mathbf{L}^{-H}, \mathbf{C}_s\mathbf{A}^T\mathbf{L}^{-T}]\begin{bmatrix} \mathbf{L}^{-1}\mathbf{R}_x\mathbf{L}^{-H} & \mathbf{L}^{-1}\mathbf{C}_x\mathbf{L}^{-T} \\ \mathbf{L}^{-*}\mathbf{C}_x^*\mathbf{L}^{-H} & \mathbf{L}^{-*}\mathbf{R}_x^*\mathbf{L}^{-T} \end{bmatrix}^{-1}\begin{bmatrix} \mathbf{L}^{-1}\mathbf{A}\mathbf{R}_s \\ \mathbf{L}^{-*}\mathbf{A}^*\mathbf{C}_s^* \end{bmatrix} \\ &= [\mathbf{R}_s\mathbf{A}^H\mathbf{L}^{-H}, \mathbf{C}_s\mathbf{A}^T\mathbf{L}^{-T}]\begin{bmatrix} \mathbf{L}^H \\ \mathbf{L}^T \end{bmatrix}\begin{bmatrix} \mathbf{R}_x & \mathbf{C}_x \\ \mathbf{C}_x^* & \mathbf{R}_x^* \end{bmatrix}^{-1}[\mathbf{L}, \mathbf{L}^*]\begin{bmatrix} \mathbf{L}^{-1}\mathbf{A}\mathbf{R}_s \\ \mathbf{L}^{-*}\mathbf{A}^*\mathbf{C}_s^* \end{bmatrix} = \mathbf{H}, \end{aligned}$$

using partitioned inverse identities (see, e.g., [32, Prop. 14.11]).

Consequently the term $(\partial\mathbf{a}/\partial\boldsymbol{\theta})^H(\mathbf{H}^T \otimes \boldsymbol{\Pi}_A^{\perp})(\partial\mathbf{a}/\partial\boldsymbol{\theta})$ in the expressions (32) and (34)–(36) becomes

$$\begin{aligned} &\left(\frac{\partial \mathbf{a}_L}{\partial \boldsymbol{\theta}}\right)^H\left(\mathbf{H}^T \otimes \boldsymbol{\Pi}_{A_L}^{\perp}\right)\left(\frac{\partial \mathbf{a}_L}{\partial \boldsymbol{\theta}}\right) \\ &= \left(\frac{\partial \mathbf{a}}{\partial \boldsymbol{\theta}}\right)^H \underbrace{(\mathbf{I} \otimes \mathbf{L}^{-H})(\mathbf{H}^T \otimes \boldsymbol{\Pi}_{A_L}^{\perp})(\mathbf{I} \otimes \mathbf{L}^{-1})}_{\mathbf{H}^T \otimes (\mathbf{L}^{-H}\boldsymbol{\Pi}_{A_L}^{\perp}\mathbf{L}^{-1})}\left(\frac{\partial \mathbf{a}}{\partial \boldsymbol{\theta}}\right), \end{aligned}$$

with

$$\mathbf{L}^{-H}\boldsymbol{\Pi}_{A_L}^{\perp}\mathbf{L}^{-1} = \boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1}\mathbf{A}(\mathbf{A}^H\boldsymbol{\Sigma}^{-1}\mathbf{A})^{-1}\mathbf{A}^H\boldsymbol{\Sigma}^{-1} \stackrel{\text{def}}{=} \boldsymbol{\Pi}_{\Sigma}. \quad \square$$

References

[1] T. Adali, P.J. Schreier, L.L. Scharf, Complex-valued signal processing: the proper way to deal with impropriety, IEEE Trans. Signal Process. 59 (November (11)) (2011) 5101–5125.

[2] E. Ollila, V. Koivunen, H.V. Poor, Complex-valued signal processing - essential models, tools and statistics, in: Proceedings of Information Theory and Applications Workshop, La Jolla, CA, 2011.

[3] A. Van Den Bos, Estimation of complex parameters, in: The 10th IFAC Symposium, vol. 3, July 1994, pp. 495–499.
[4] T. Adalı, Complex-valued adaptive signal processing, in: Adaptive Signal Processing—Next-Generation Solutions, John Wiley and Sons, New York, 2010.
[5] P. Schreier, L.L. Scharf, Statistical Signal Processing of Complex-Valued Data—The Theory of Improper and Noncircular Signals, Cambridge University Press, New York, 2010.
[6] J.P. Delmas, Asymptotic performance of second-order algorithms, IEEE Trans. Signal Process. 50 (January (1)) (2002) 49–57.
[7] J.P. Delmas, Performance bounds and statistical Analysis of DOA estimation, in: Academic Press Library in Signal Processing, vol. 3, 1st edition, Array and Statistical Signal Processing, Elsevier, Oxford, UK, 2013.
[8] J.R. Magnus, H. Neudecker, Matrix Differential Calculus with Applications in Statistics and Econometrics, Revised edition, John Wiley and Sons, UK, 1999.
[9] J.P. Delmas, H. Abeida, Asymptotic distribution of circularity coefficients estimate of complex random variables, Signal Process. 89 (December) (2009) 2311–2698.
[10] J. Eriksson, E. Ollila, V. Koivunen, Essential statistics and tools for complex random variables, IEEE Trans. Signal Process. 58 (October (10)) (2010) 5400–5408.
[11] K. Kreutz-Delgado, The Complex Gradient Operator and the CR-calculus, Cornell University Library, New York, 2007.
[12] E.L. Lehmann, Elements of Large-Sample Theory, Springer-Verlag, New-York, 1999.
[13] G.B. Giannakis, Cumulants: a powerful tool in signal processing, Proc. IEEE 75 (September (9)) (1987) 1333–1334.
[14] J.P. Delmas, Y. Meurisse, Extension of the matrix Bartlett's formula to the third and fourth order and to noisy linear models with application to parameter estimation, IEEE Trans. Signal Process. 53 (August (8)) (2005) 2765–2776.
[15] B. Porat, B. Friedlander, Performance analysis of parameter estimation algorithms based on high-order moments, Int. J. Adapt. Control Signal Process. 3 (1989) 191–229.
[16] P. Stoica, B. Friedlander, T. Söderström, An approximate maximum approach to ARMA spectral estimation, in: Proceedings of the Decision and Control, Fort Lauderdale, 1985.
[17] J.P. Delmas, Asymptotically minimum variance second-order estimation for noncircular signals with application to DOA estimation, IEEE Trans. Signal Process. 52 (May (5)) (2004) 1235–1241.
[18] A. Van Den Bos, A Cramer–Rao lower bound for complex parameters, IEEE Trans. Signal Process. 42 (10) (1994) 2859.
[19] E. Ollila, V. Koivunen, J. Eriksson, On the Cramer-Rao bound for the constrained and unconstrained complex parameters, in: Proceedings of IEEE Sensor Array and Multichannel Signal Processing Workshop, Darmstadt, Germany, July 2008.
[20] B. Loesch, B. Yang, Cramer-Rao bound for circular and noncircular complex independent component analysis, IEEE Trans. Signal Process. 61 (January (2)) (2013) 365–379.
[21] J.P. Delmas, H. Abeida, Stochastic Cramer–Rao bound for noncircular signals with application to DOA estimation, IEEE Trans. Signal Process. 52 (November (11)) (2004) 3192–3199.
[22] P. Whittle, The analysis of multiple stationary time series, J. R. Stat. Soc. 15 (1953) 125–139.
[23] J.P. Delmas, Y. Meurisse, On the Cramer Rao bound and maximum likelihood in passive time delay estimation for complex signals, in: Proceedings of ICASSP, Kyoto, Japan, March 2012.
[24] P. Stoica, A.G. Larsson, A.B. Gershman, The stochastic CRB for array processing: a textbook derivation, IEEE Signal Process. Lett. 8 (May) (2001) 148–150.
[25] M. Jansson, B. Goransson, B. Ottersten, A subspace method for direction of arrival estimation of uncorrelated emitter signals, IEEE Trans. Signal Process. 47 (April (4)) (1999) 945–956.
[26] J.F. Cardoso, E. Moulines, Invariance of subspace based estimator, IEEE Trans. Signal Process. 48 (September (9)) (2000) 2495–2505.
[27] H. Abeida, J.P. Delmas, Efficiency of subspace-based DOA estimators, Signal Process. 87 (September (9)) (2007) 2075–2084.
[28] J.P. Delmas, P. Comon, Y. Meurisse, Performance limits of alphabet diversities for FIR SISO channel identification, IEEE Trans. Signal Process. 57 (January (1)) (2009) 73–82.
[29] D. Darsena, G. Gelli, L. Paura, F. Verde, Subspace-based blind channel identification of SISO-FIR systems with improper random inputs, Signal Process. 84 (2004) 2021–2039.
[30] R.J. Serfling, Approximation Theorems of Mathematical Statistics, John Wiley and Sons, New York, 1980.
[31] P. Stoica, A. Nehorai, Performance study of conditional and unconditional direction of arrival estimation, IEEE Trans. ASSP 38 (October) (1990) 1783–1795.
[32] G.A.F. Seber, A Matrix Handbook for Statisticians, Wiley Series in Probability and Statistics, NJ, 2008.