Cramér-Rao Bound for Circular and Noncircular Complex Independent Component Analysis

Benedikt Loesch*, *Student Member, IEEE,* and Bin Yang, *Senior Member, IEEE,*

***Abstract*—Despite an increased interest in complex independent component analysis (ICA) during the last two decades, a closed form expression for the Cramér-Rao bound (CRB) for the demixing matrix is not known yet. In this paper, we fill this gap by deriving a closed-form expression for the CRB of the demixing matrix for instantaneous noncircular complex ICA. It contains the CRB for circular complex ICA and noncircular complex Gaussian ICA as two special cases. We also study the CRB numerically for the family of noncircular complex generalized Gaussian distributions and compare it to simulation results of two ICA estimators. Furthermore, we show how to extend the CRB to the case where the source signals are not temporally independent and identically distributed.**

***Index Terms*—Independent component analysis, Cramér-Rao bound, noncircular complex, generalized Gaussian distribution**

I. Introduction

Independent Component Analysis (ICA) is a signal processing method (see [1]–[3]) to extract unobservable source signals or independent components from their observable linear mixtures. We assume an instantaneous complex linear square noiseless mixing model

$$\mathbf{x} = \mathbf{A}\mathbf{s} \tag{1}$$

where $\mathbf{x} \in \mathbb{C}^N$ are N linear combinations of the N source signals $\mathbf{s} \in \mathbb{C}^N$. We make the following assumptions:

A1) The mixing matrix $\mathbf{A} \in \mathbb{C}^{N\times N}$ is deterministic and invertible.

A2) $\mathbf{s} = [s_1, \cdots, s_N]^T \in \mathbb{C}^N$ are N independent random variables with zero mean, unit variance $\mathrm{E}\left[|s_i|^2\right] = 1$ and second-order noncircularity index $\gamma_i = E[s_i^2] \in [0, 1]$ (after scaling the columns of $\mathbf{A}$ suitably). Since $\gamma_i \in \mathbb{R}$, the real and imaginary part of s_i are uncorrelated. $\gamma_i \neq 0$ if and only if the variances of the real and imaginary part of s_i differ. The probability density functions (pdfs) $p_i(s_i)$ of different source signals s_i can be identical or different. $p_i(s_i)$ is continuously differentiable with respect to s_i and s_i^* in the sense of Wirtinger derivatives [4] which will be shortly reviewed in Sec. II. All required expectations exist.

The task of ICA is to demix the signals $\mathbf{x}$ by a linear demixing matrix $\mathbf{W} \in \mathbb{C}^{N\times N}$

$$\mathbf{y} = \mathbf{W}\mathbf{x} = \mathbf{W}\mathbf{A}\mathbf{s} \tag{2}$$

such that $\mathbf{y}$ is "as close to $\mathbf{s}$" as possible according to some metric. General conditions regarding identifiability, uniqueness and separability can be found in [5]. The ideal solution for $\mathbf{W}$ is $\mathbf{A}^{-1}$, neglecting scaling, phase and permutation ambiguity [5]. If we know the pdfs $p_i(s_i)$ perfectly, there is no scaling ambiguity. Due to the "working" assumption $\gamma_i \in [0, 1]$ (see A2), there is no phase ambiguity for noncircular sources ($\gamma_i > 0$). A phase ambiguity occurs only for circular sources ($\gamma_i = 0$). Noncircular sources which do not comply with the assumption $\gamma_i \in [0, 1]$ would be reconstructed subject to a phase shift such that $\gamma_i \in [0, 1]$. [6]–[8] provide a neural network view for the theory of complex ICA and illustrate a number of applications. While [6]–[8] and many other publications focus on circular complex signals (as traditionally assumed in signal processing), [9] provides a good overview of applications with noncircular complex signals and discusses how to properly deal with noncircularity.

In general, a complex source signal s can be described by the following statistical properties:

- non-Gaussianity
- noncircularity
- nonwhiteness, i.e. $s(t_1)$ and $s(t_2)$ are dependent for some different time instants $t_1 \neq t_2$
- nonstationarity, i.e. the statistical properties of $s(t)$ change over time

In this paper, we will mainly focus on noncircular complex source signals with independent and identically distributed (iid) time samples. However, we also show how to extend our results to temporally non-iid sources, i.e. to incorporate nonstationarity and nonwhiteness of the sources.

Two temporally iid sources can be separated by ICA

- if at least one of the two sources is non-Gaussian or
- if both sources are Gaussian but differ in noncircularity [5].

For the performance analysis of ICA algorithms, it is useful to have a lower bound for the covariance matrix of estimators for the demixing matrix $\mathbf{W}$. The Cramér-Rao bound (CRB) is a lower bound on the covariance matrix of any unbiased estimator of a parameter vector. A closed-form expression for the CRB of the demixing matrix for real instantaneous ICA has been derived recently in [10], [11]. However, in many practical applications such as audio processing in frequency domain or telecommunication, the signals are complex and hence we need a complex ICA algorithm. Although many different algorithms for complex ICA have been proposed [7], [12]–[20], the CRB for the complex demixing matrix has not been derived yet. [21] provides a performance analysis

B. Loesch and B. Yang are with the Institute of Signal Processing and System Theory at the Universität Stuttgart, Pfaffenwaldring 47, 70550 Stuttgart, Germany. E-mail: benedikt.loesch@iss.uni-stuttgart.de; bin.yang@iss.uni-stuttgart.de;
EDICS: MLR-ICAN, MLR-SSEP, SSP-SSEP, SSP-PERF, MLR-PERF

for the Strong Uncorrelating Transform (SUT) in terms of the interference-to-signal ratio matrix. However, since the SUT uses only second-order statistics, the results from [21] do not apply for ICA algorithms exploiting also the non-Gaussianity of the sources. As discussed in [22] and [9], many ICA approaches exploiting non-Gaussianity of the sources are intimately related and can be studied under the umbrella of a maximum likelihood framework. In [23], we derived the CRB for the non-Gaussian circular complex case. In this paper, we extend the derivation to the noncircular complex case and provide a more thorough performance study.

The paper is organized as follows: In Sec. II we briefly review complex random vectors, complex gradient and the CRB for a complex parameter vector. In Sec. III we derive the CRB for the complex demixing matrix $\mathbf{W} = \mathbf{A}^{-1}$ and Sec. IV discusses the circular complex case and noncircular complex Gaussian case as two special cases of the CRB derived in Sec. III. Sec. V considers again the general noncircular complex case and shows numerical results for the CRB as well as simulation results of two ICA algorithms to verify the CRB. Sec. VI shows how to extend our results to temporally non-iid sources. Finally, concluding remarks are given in Sect. VII.

II. Complex notations

A. Complex random vector

Let $\mathbf{x} = \mathbf{x}_R + j\mathbf{x}_I \in \mathbb{C}^N$ be a complex random vector with a corresponding probability density function (pdf) defined as the pdf $\tilde{p}(\mathbf{x}_R, \mathbf{x}_I)$ of the real part $\mathbf{x}_R$ and imaginary part $\mathbf{x}_I$ of $\mathbf{x}$. Since $\mathbf{x}_R = \frac{\mathbf{x}+\mathbf{x}^*}{2}$ and $\mathbf{x}_I = \frac{\mathbf{x}-\mathbf{x}^*}{2j}$, we can rewrite the pdf $\tilde{p}(\mathbf{x}_R, \mathbf{x}_I)$ as a function of $\mathbf{x}$ and $\mathbf{x}^*$, i.e. $\tilde{p}(\mathbf{x}_R, \mathbf{x}_I) = p(\mathbf{x}, \mathbf{x}^*)$. In the following, we will use $p(\mathbf{x})$ as a short notation for $p(\mathbf{x}, \mathbf{x}^*)$. The covariance matrix of $\mathbf{x}$ is

$$\operatorname{cov}(\mathbf{x}) = \mathrm{E}\left[(\mathbf{x} - \mathrm{E}[\mathbf{x}])(\mathbf{x} - \mathrm{E}[\mathbf{x}])^H\right]. \tag{3}$$

The pseudo-covariance matrix of $\mathbf{x}$ is

$$\operatorname{pcov}(\mathbf{x}) = \mathrm{E}\left[(\mathbf{x} - \mathrm{E}[\mathbf{x}])(\mathbf{x} - \mathrm{E}[\mathbf{x}])^T\right]. \tag{4}$$

The augmented covariance matrix of $\mathbf{x}$ is the covariance matrix of the augmented vector $\underline{\mathbf{x}} = \begin{bmatrix}\mathbf{x}^T & \mathbf{x}^H\end{bmatrix}^T$:

$$\operatorname{cov}(\underline{\mathbf{x}}) = \begin{bmatrix}\operatorname{cov}(\mathbf{x}) & \operatorname{pcov}(\mathbf{x}) \\ \operatorname{pcov}(\mathbf{x})^* & \operatorname{cov}(\mathbf{x})^*\end{bmatrix}. \tag{5}$$

$\mathbf{x}$ is called circular if $p(\mathbf{x}e^{j\alpha}) = p(\mathbf{x})\ \forall \alpha \in \mathbb{R}$. Otherwise it is called noncircular. Actually, for a random variable s, the circularity definition $p(se^{j\alpha}) = p(s)\ \forall \alpha \in \mathbb{R}$ is much stronger than the second-order circularity given by $\gamma = \mathrm{E}\left[s^2\right] = 0$. Indeed there exist noncircular complex random variables with $\gamma = 0$. For simplicity, however, we use the second-order noncircularity index $\gamma = \mathrm{E}\left[s^2\right]$ to quantify noncircularity in the remainder of the paper.

B. Complex gradient

Let a complex *column* parameter vector $\boldsymbol{\theta} = \boldsymbol{\theta}_R + j\boldsymbol{\theta}_I \in \mathbb{C}^M$, its real and imaginary part $\boldsymbol{\theta}_R, \boldsymbol{\theta}_I \in \mathbb{R}^M$, and a real scalar cost function $f(\boldsymbol{\theta}, \boldsymbol{\theta}^*) = \tilde{f}(\boldsymbol{\theta}_R, \boldsymbol{\theta}_I) \in \mathbb{R}$ be given. For ease of notation, we will also use the simplified notation $f(\boldsymbol{\theta})$ instead of $f(\boldsymbol{\theta}, \boldsymbol{\theta}^*)$. Instead of calculating the derivatives of $\tilde{f}(\cdot)$ with respect to $\boldsymbol{\theta}_R$ and $\boldsymbol{\theta}_I$, the Wirtinger calculus computes the partial derivatives of $f(\boldsymbol{\theta}, \boldsymbol{\theta}^*)$ with respect to $\boldsymbol{\theta}$ and $\boldsymbol{\theta}^*$, treating $\boldsymbol{\theta}$ and $\boldsymbol{\theta}^*$ as two independent variables [24], [25]. The partial derivatives of $f(\cdot)$ with respect to $\boldsymbol{\theta}$ and $\boldsymbol{\theta}^*$ and the complex gradient vectors $\nabla_{\boldsymbol{\theta}} f$ and $\nabla_{\boldsymbol{\theta}^*} f$ are

$$\nabla_{\boldsymbol{\theta}} f = \frac{\partial f}{\partial \boldsymbol{\theta}} = \frac{1}{2}\left(\frac{\partial \tilde{f}}{\partial \boldsymbol{\theta}_R} - j\frac{\partial \tilde{f}}{\partial \boldsymbol{\theta}_I}\right) \in \mathbb{C}^M,$$
$$\nabla_{\boldsymbol{\theta}^*} f = \frac{\partial f}{\partial \boldsymbol{\theta}^*} = \frac{1}{2}\left(\frac{\partial \tilde{f}}{\partial \boldsymbol{\theta}_R} + j\frac{\partial \tilde{f}}{\partial \boldsymbol{\theta}_I}\right) \in \mathbb{C}^M. \tag{6}$$

The stationary point of $f(\cdot)$ and $\tilde{f}(\cdot)$ is given by $\frac{\partial \tilde{f}}{\partial \boldsymbol{\theta}_R} = \mathbf{0}$ and $\frac{\partial \tilde{f}}{\partial \boldsymbol{\theta}_I} = \mathbf{0}$ or $\frac{\partial f}{\partial \boldsymbol{\theta}} = \mathbf{0}$ or $\frac{\partial f}{\partial \boldsymbol{\theta}^*} = \mathbf{0}$. The direction of steepest descent of a real function $f(\boldsymbol{\theta}, \boldsymbol{\theta}^*)$ is given by $-\frac{\partial f}{\partial \boldsymbol{\theta}^*}$ and not $-\frac{\partial f}{\partial \boldsymbol{\theta}}$ [26]. Note that $-\frac{\partial f}{\partial \boldsymbol{\theta}^*}$ is the direction of steepest descent for $\boldsymbol{\theta}$ and not for $\boldsymbol{\theta}^*$.

As long as the real and imaginary part of a complex function $g(\boldsymbol{\theta}, \boldsymbol{\theta}^*) = g_R(\boldsymbol{\theta}_R, \boldsymbol{\theta}_I) + jg_I(\boldsymbol{\theta}_R, \boldsymbol{\theta}_I)$ are differentiable, the Wirtinger derivatives $\frac{\partial g}{\partial \boldsymbol{\theta}} = \frac{\partial g_R}{\partial \boldsymbol{\theta}} + j\frac{\partial g_I}{\partial \boldsymbol{\theta}}$ and $\frac{\partial g}{\partial \boldsymbol{\theta}^*} = \frac{\partial g_R}{\partial \boldsymbol{\theta}^*} + j\frac{\partial g_I}{\partial \boldsymbol{\theta}^*}$ also exist [27].

C. Cramér-Rao bound for a complex parameter vector

Assume that L complex observations of $\mathbf{x}$ are iid with the pdf $p(\mathbf{x}; \boldsymbol{\theta})$ where $\boldsymbol{\theta}$ is an N-dimensional complex parameter vector. In principle, it would be possible to derive the CRB for complex parameter $\boldsymbol{\theta} = \boldsymbol{\theta}_R + j\boldsymbol{\theta}_I$ by considering the real CRB of the $2N$-dimensional real composite vector $\bar{\boldsymbol{\theta}} = \begin{bmatrix}\boldsymbol{\theta}_R^T & \boldsymbol{\theta}_I^T\end{bmatrix}^T$:

$$\operatorname{cov}(\bar{\boldsymbol{\theta}}) = \begin{bmatrix}\operatorname{cov}(\boldsymbol{\theta}_R) & \operatorname{cov}(\boldsymbol{\theta}_R, \boldsymbol{\theta}_I) \\ \operatorname{cov}(\boldsymbol{\theta}_I, \boldsymbol{\theta}_R) & \operatorname{cov}(\boldsymbol{\theta}_I)\end{bmatrix} \geq L^{-1}\mathbf{J}_{\bar{\boldsymbol{\theta}}}^{-1}, \tag{7}$$

where $\operatorname{cov}(\mathbf{x}, \mathbf{y}) = \mathrm{E}\left[(\mathbf{x} - \mathrm{E}[\mathbf{x}])(\mathbf{y} - \mathrm{E}[\mathbf{y}])^T\right]$ is the covariance of $\mathbf{x}$ and $\mathbf{y}$, $\mathbf{J}_{\bar{\boldsymbol{\theta}}} = \mathrm{E}\left[\left\{\nabla_{\bar{\boldsymbol{\theta}}} \ln p(\mathbf{x}; \bar{\boldsymbol{\theta}})\right\}\left\{\nabla_{\bar{\boldsymbol{\theta}}} \ln p(\mathbf{x}; \bar{\boldsymbol{\theta}})\right\}^T\right]$ is the real Fisher Information matrix (FIM) and $\nabla_{\bar{\boldsymbol{\theta}}} \ln p(\mathbf{x}; \bar{\boldsymbol{\theta}})$ is the real gradient vector of $\ln p(\mathbf{x}; \bar{\boldsymbol{\theta}})$.

However, it is often more convenient to directly work with the complex CRB introduced in this section: The complex FIM of $\boldsymbol{\theta}$ is defined as

$$\mathcal{J}_{\boldsymbol{\theta}} = \begin{bmatrix}\mathcal{I}_{\boldsymbol{\theta}} & \mathcal{P}_{\boldsymbol{\theta}} \\ \mathcal{P}_{\boldsymbol{\theta}}^* & \mathcal{I}_{\boldsymbol{\theta}}^*\end{bmatrix}, \tag{8}$$

where $\mathcal{I}_{\boldsymbol{\theta}} = \mathrm{E}\left[\left\{\nabla_{\boldsymbol{\theta}^*} \ln p(\mathbf{x}; \boldsymbol{\theta})\right\}\left\{\nabla_{\boldsymbol{\theta}^*} \ln p(\mathbf{x}; \boldsymbol{\theta})\right\}^H\right]$ and $\mathcal{P}_{\boldsymbol{\theta}} = \mathrm{E}\left[\left\{\nabla_{\boldsymbol{\theta}^*} \ln p(\mathbf{x}; \boldsymbol{\theta})\right\}\left\{\nabla_{\boldsymbol{\theta}^*} \ln p(\mathbf{x}; \boldsymbol{\theta})\right\}^T\right]$ are called the information matrix and pseudo-information matrix.

The inverse of the FIM of $\boldsymbol{\theta}$ gives, under some regularity conditions, a lower bound for the augmented covariance matrix of an unbiased estimator $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$ [25], [28]

$$\begin{bmatrix}\operatorname{cov}(\hat{\boldsymbol{\theta}}) & \operatorname{pcov}(\hat{\boldsymbol{\theta}}) \\ \operatorname{pcov}(\hat{\boldsymbol{\theta}})^* & \operatorname{cov}(\hat{\boldsymbol{\theta}})^*\end{bmatrix} \geq (L\mathcal{J}_{\boldsymbol{\theta}})^{-1} = L^{-1}\begin{bmatrix}\mathcal{I}_{\boldsymbol{\theta}} & \mathcal{P}_{\boldsymbol{\theta}} \\ \mathcal{P}_{\boldsymbol{\theta}}^* & \mathcal{I}_{\boldsymbol{\theta}}^*\end{bmatrix}^{-1}. \tag{9}$$

Note that the complex CRB (9) can be transformed to the corresponding real CRB (7) by using the transform $\mathbf{J}_{\bar{\boldsymbol{\theta}}}^{-1} = \frac{1}{2}\mathbf{T}\mathcal{J}_{\boldsymbol{\theta}}^{-1}\mathbf{T}^{-1}$ [28], where $\mathbf{T} = \frac{1}{2}\begin{bmatrix}\mathbf{I} & \mathbf{I} \\ -j\mathbf{I} & j\mathbf{I}\end{bmatrix}$ is a $2N \times 2N$ matrix and $\mathbf{I}$ is the $N \times N$ identity matrix.

By using the block matrix inversion lemma [29], we get from (9)

$$\begin{bmatrix} \text{cov}(\hat{\boldsymbol{\theta}}) & \text{pcov}(\hat{\boldsymbol{\theta}}) \\ \text{pcov}(\hat{\boldsymbol{\theta}})^* & \text{cov}(\hat{\boldsymbol{\theta}})^* \end{bmatrix} \geq L^{-1} \begin{bmatrix} \mathbf{R}_{\boldsymbol{\theta}}^{-1} & -\mathbf{R}_{\boldsymbol{\theta}}^{-1}\mathbf{Q}_{\boldsymbol{\theta}} \\ -\mathbf{Q}_{\boldsymbol{\theta}}^H \mathbf{R}_{\boldsymbol{\theta}}^{-1} & \mathbf{R}_{\boldsymbol{\theta}}^{-*} \end{bmatrix} \quad (10)$$

with $\mathbf{R}_{\boldsymbol{\theta}} = \mathcal{I}_{\boldsymbol{\theta}} - \mathcal{P}_{\boldsymbol{\theta}}\mathcal{I}_{\boldsymbol{\theta}}^{-*}\mathcal{P}_{\boldsymbol{\theta}}^*$ and $\mathbf{Q}_{\boldsymbol{\theta}} = \mathcal{P}_{\boldsymbol{\theta}}\mathcal{I}_{\boldsymbol{\theta}}^{-*}$. $\mathbf{A}^{-*}$ is a short notation for $\left(\mathbf{A}^{-1}\right)^* = (\mathbf{A}^*)^{-1}$. Often we are interested in the bound for $\text{cov}(\hat{\boldsymbol{\theta}})$ only, which can be obtained from (10) as

$$\text{cov}(\hat{\boldsymbol{\theta}}) \geq L^{-1}\mathbf{R}_{\boldsymbol{\theta}}^{-1} = L^{-1}(\mathcal{I}_{\boldsymbol{\theta}} - \mathcal{P}_{\boldsymbol{\theta}}\mathcal{I}_{\boldsymbol{\theta}}^{-*}\mathcal{P}_{\boldsymbol{\theta}}^*)^{-1}. \quad (11)$$

Note that (11) gives a bound solely on the covariance matrix of an unbiased estimator. If an estimator reaches that bound, i.e. $\text{cov}(\hat{\boldsymbol{\theta}}) = L^{-1}\mathbf{R}_{\boldsymbol{\theta}}^{-1}$, it does not imply that it also reaches the general CRB defined in (9). Only if the pseudo-information matrix $\mathcal{P}_{\boldsymbol{\theta}}$ vanishes, $\text{cov}(\hat{\boldsymbol{\theta}}) = L^{-1}\mathbf{R}_{\boldsymbol{\theta}}^{-1}$ implies that $\hat{\boldsymbol{\theta}}$ reaches the CRB (9).

Sometimes, we are interested in introducing constraints on some or all of the complex parameters. The constrained CRB can be derived by following the steps in [28] or [30]. If the unconstrained Fisher information matrix is singular, we have to use the constrained CRB from [30] which is briefly reviewed in Appendix D.

III. Derivation of Cramér-Rao bound

We form the parameter vector

$$\boldsymbol{\theta} = \text{vec}(\mathbf{W}^T) = [\mathbf{w}_1^T, \cdots, \mathbf{w}_N^T]^T \in \mathbb{C}^{N^2} \quad (12)$$

where $\mathbf{w}_i^T$ denotes the i-th row vector of $\mathbf{W}$. The $\text{vec}(\cdot)$ operator stacks the columns of its argument into one long column vector. Given the pdfs $p_i(s_i)$ of the complex source signals s_i and the complex linear transform $\mathbf{x} = \mathbf{A}\mathbf{s}$, it is easy to derive the pdf of $\mathbf{x}$ as $p(\mathbf{x}; \boldsymbol{\theta}) = |\det(\mathbf{W})|^2 \prod_{i=1}^N p_i(\mathbf{w}_i^T\mathbf{x})$. Here, in the derivation of the CRB, $\mathbf{W}$ is a short notation for $\mathbf{A}^{-1}$ and not the demixing matrix which would contain permutation, scaling and phase ambiguity. By using matrix derivatives [22], [24], [31], we obtain

$$\frac{\partial}{\partial \mathbf{W}^H} \ln p(\mathbf{x}; \boldsymbol{\theta}) = \mathbf{A}^* - \mathbf{x}^*\boldsymbol{\varphi}^T(\mathbf{W}\mathbf{x}) = \mathbf{A}^*(\mathbf{I} - \mathbf{s}\boldsymbol{\varphi}^H(\mathbf{s}))^* \quad (13)$$

where $\boldsymbol{\varphi}(\mathbf{s}) = [\varphi_1(s_1), \cdots, \varphi_N(s_N)]^T$ and $\varphi_i(s_i)$ is defined as

$$\varphi_i(s_i) = -\frac{\partial}{\partial s_i^*} \ln p_i(s_i) = -\frac{1}{2}\frac{1}{p_i(s_i)}\left[\frac{\partial p_i(s_i)}{\partial s_{i,R}} + j\frac{\partial p_i(s_i)}{\partial s_{i,I}}\right]. \quad (14)$$

Since $\boldsymbol{\theta} = \text{vec}(\mathbf{W}^T)$, we get $\nabla_{\boldsymbol{\theta}^*} \ln p(\mathbf{x}; \boldsymbol{\theta}) = \text{vec}\left(\frac{\partial}{\partial \mathbf{W}^H} \ln p(\mathbf{x}; \boldsymbol{\theta})\right) = \left[(\mathbf{I} \otimes \mathbf{A})\,\text{vec}\left(\mathbf{I} - \mathbf{s}\boldsymbol{\varphi}(\mathbf{s})^H\right)\right]^*$, where $\mathbf{A} \otimes \mathbf{B} = [a_{ij}\mathbf{B}]$ denotes the Kronecker product of $\mathbf{A}$ and $\mathbf{B}$. Hence, the information and pseudo-information matrix in (8) become

$$\begin{aligned} \mathcal{I}_{\boldsymbol{\theta}} &= \left((\mathbf{I} \otimes \mathbf{A})\,\text{E}\left[\text{vec}\{\mathbf{I} - \mathbf{s}\boldsymbol{\varphi}^H(\mathbf{s})\}\,\text{vec}\{\mathbf{I} - \mathbf{s}\boldsymbol{\varphi}^H(\mathbf{s})\}^H\right] \right. \\ &\quad \left. \cdot(\mathbf{I} \otimes \mathbf{A}^H)\right)^* \\ &= \left((\mathbf{I} \otimes \mathbf{A})\mathbf{M}_1(\mathbf{I} \otimes \mathbf{A}^H)\right)^*, \quad (15) \\ \mathcal{P}_{\boldsymbol{\theta}} &= \left((\mathbf{I} \otimes \mathbf{A})\,\text{E}\left[\text{vec}\{\mathbf{I} - \mathbf{s}\boldsymbol{\varphi}^H(\mathbf{s})\}\,\text{vec}\{\mathbf{I} - \mathbf{s}\boldsymbol{\varphi}^H(\mathbf{s})\}^T\right] \right. \\ &\quad \left. \cdot(\mathbf{I} \otimes \mathbf{A}^T)\right)^* \\ &= \left((\mathbf{I} \otimes \mathbf{A})\mathbf{M}_2(\mathbf{I} \otimes \mathbf{A}^T)\right)^*, \quad (16) \end{aligned}$$

where

$$\begin{aligned} \mathbf{M}_1 &= \text{E}\left[\text{vec}\{\mathbf{I} - \mathbf{s}\boldsymbol{\varphi}^H(\mathbf{s})\}\,\text{vec}\{\mathbf{I} - \mathbf{s}\boldsymbol{\varphi}^H(\mathbf{s})\}^H\right] \\ \text{and } \mathbf{M}_2 &= \text{E}\left[\text{vec}\{\mathbf{I} - \mathbf{s}\boldsymbol{\varphi}^H(\mathbf{s})\}\,\text{vec}\{\mathbf{I} - \mathbf{s}\boldsymbol{\varphi}^H(\mathbf{s})\}^T\right]. \quad (17) \end{aligned}$$

A. *Induced CRB for the gain matrix* $\mathbf{G} = \mathbf{W}\mathbf{A}$

Since the so-called gain matrix $\mathbf{G} = \mathbf{W}\mathbf{A}$ is a linear function of $\mathbf{W}$, the CRB for $\mathbf{W}$ “induces“ a bound for $\mathbf{G}$. For simplicity, we first derive this induced CRB (iCRB) for $\mathbf{G} = \mathbf{W}\mathbf{A} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}$ which is independent of the mixing matrix $\mathbf{A}$. Later we will obtain the CRB for $\mathbf{W}$ from the iCRB for $\mathbf{G}$[1]. When $\hat{\mathbf{G}} = \hat{\mathbf{W}}\mathbf{A}$ denotes the estimated gain matrix, the diagonal elements $\hat{G}_{ii}$ should be close to 1. They reflect how well we can estimate the power of each source signal. The off-diagonal elements $\hat{G}_{ij}$ should be close to 0 and reflect how well we can suppress interfering components. We define the corresponding stacked parameter vector

$$\begin{aligned} \boldsymbol{\vartheta} &= \text{vec}(\mathbf{G}^T) = \text{vec}(\mathbf{A}^T\mathbf{W}^T) = (\mathbf{I} \otimes \mathbf{A}^T)\,\text{vec}(\mathbf{W}^T) \\ &= (\mathbf{I} \otimes \mathbf{A}^T)\boldsymbol{\theta}. \quad (18) \end{aligned}$$

The covariance matrix of $\hat{\boldsymbol{\vartheta}} = \text{vec}((\hat{\mathbf{W}}\mathbf{A})^T)$ is given by $\text{cov}(\hat{\boldsymbol{\vartheta}}) = (\mathbf{I} \otimes \mathbf{A}^T)\,\text{cov}(\hat{\boldsymbol{\theta}})(\mathbf{I} \otimes \mathbf{A}^*)$ where $\hat{\boldsymbol{\theta}} = \text{vec}(\hat{\mathbf{W}}^T)$. By combining (11) with (15) and (16), we get

$$\begin{aligned} \text{cov}(\hat{\boldsymbol{\vartheta}}) &\geq L^{-1}(\mathbf{I} \otimes \mathbf{A}^T)(\mathcal{I}_{\boldsymbol{\theta}} - \mathcal{P}_{\boldsymbol{\theta}}\mathcal{I}_{\boldsymbol{\theta}}^{-*}\mathcal{P}_{\boldsymbol{\theta}}^*)^{-1}(\mathbf{I} \otimes \mathbf{A}^*) \\ &= L^{-1}\mathbf{R}_{\boldsymbol{\vartheta}}^{-1} \quad (19) \end{aligned}$$

with

$$\mathbf{R}_{\boldsymbol{\vartheta}} = (\mathbf{M}_1 - \mathbf{M}_2\mathbf{M}_1^{-*}\mathbf{M}_2^*)^*. \quad (20)$$

As shown in Appendix C, $\mathbf{R}_{\boldsymbol{\vartheta}}$ can be calculated as

$$\begin{aligned} \mathbf{R}_{\boldsymbol{\vartheta}} = &\sum_{i=1}^N d_i \mathbf{L}_{ii} \otimes \mathbf{L}_{ii} + \sum_{i=1}^N \sum_{\substack{j=1 \\ j\neq i}}^N a_{ij}\mathbf{L}_{ii} \otimes \mathbf{L}_{jj} \\ &+ \sum_{i=1}^N \sum_{\substack{j=1 \\ j\neq i}}^N b_{ij}\mathbf{L}_{ij} \otimes \mathbf{L}_{ji} \quad (21) \end{aligned}$$

where $d_i = \frac{(\eta_i - 1)^2 - |\beta_i - 1|^2}{\eta_i - 1} \in \mathbb{R}$, $a_{ij} = \kappa_i - \frac{|\gamma_j \xi_i|^2}{\kappa_i} - \frac{1}{\kappa_j} \in \mathbb{R}$ and $b_{ij} = -\left(\frac{\gamma_j^* \xi_i^*}{\kappa_i} + \frac{\gamma_i \xi_j}{\kappa_j}\right) = b_{ji}^* \in \mathbb{C}$. $\mathbf{L}_{ij}$ in (21) denotes an $N \times N$ matrix with a 1 at the (i, j) position and 0's elsewhere. The parameters η_i, κ_i, β_i, ξ_i and γ_j are defined as

$$\eta_i = \text{E}\left[|s_i|^2 |\varphi_i(s_i)|^2\right] > 1, \quad (22)$$

$$\kappa_i = \text{E}\left[|\varphi_i(s_i)|^2\right] \geq 1, \quad (23)$$

$$\beta_i = \text{E}\left[s_i^2(\varphi_i^*(s_i))^2\right] \in \mathbb{C}, \quad (24)$$

[1]Some authors [32]–[34] prefer the so-called interference-to-source ratio (ISR) matrix whose elements ISR_{ij} are defined (for $i \neq j$ and unit variance sources) as $\text{ISR}_{ij} = \text{E}\left[\frac{|G_{ij}|^2}{|G_{ii}|^2}\right]$, where G_{ii} denotes the diagonal elements and G_{ij} the off-diagonal elements of $\mathbf{G}$. To compute ISR_{ij}, usually $G_{ii} \approx 1$ (i.e. $\text{var}(G_{ii}) \ll 1$) is assumed such that $\text{ISR}_{ij} \approx \text{var}(G_{ij})$. In this paper, we do not use the ISR matrix but instead directly derive the iCRB for $\mathbf{G}$.

$$\xi_i = \mathrm{E}\left[(\varphi_i^*(s_i))^2\right] \in \mathbb{C}, \quad (25)$$

$$\gamma_j = \mathrm{E}\left[s_j^2\right] \in \mathbb{R}, \quad (26)$$

Appendix B shows some properties and other equivalent forms of these parameters. All of them depend on the pdf $p_i(s_i)$ of the i-th source signal. κ_i is a measure of non-Gaussianity and noncircularity. As shown in [19], $\kappa_i \geq 1$ with equality if and only if s_i is circular complex Gaussian. In addition, $\eta_i > 1$ (see Corollary 3). If the pdf of s_i is symmetric in the real or imaginary part of s_i, i.e. $p(-s_R, s_I) = p(s_R, s_I)$ or $p(s_R, -s_I) = p(s_R, s_I)$, β_i and ξ_i become real (see Lemma 2). γ_j is real due to assumption A2 in Sec. I.

$\mathbf{R}_{\boldsymbol{\vartheta}}$ has a special sparse structure which is illustrated in Fig. 1 for $N = 3$. The i-th diagonal element of the i-th

$$\mathbf{R}_{\boldsymbol{\vartheta}} = \begin{bmatrix} d_1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & a_{12} & 0 & b_{12} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & a_{13} & 0 & 0 & 0 & b_{13} & 0 & 0 \\ 0 & b_{21} & 0 & a_{21} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & d_2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & a_{23} & 0 & b_{23} & 0 \\ 0 & 0 & b_{31} & 0 & 0 & 0 & a_{31} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & b_{32} & 0 & a_{32} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & d_3 \end{bmatrix}$$

Fig. 1: Structure of $\mathbf{R}_{\boldsymbol{\vartheta}}$ for $N = 3$

diagonal block is $\mathbf{R}_{\boldsymbol{\vartheta}}[i,i]_{(i,i)} = d_i$. The j-th diagonal element of the i-th diagonal block is $\mathbf{R}_{\boldsymbol{\vartheta}}[i,i]_{(j,j)} = a_{ij}$. The (j,i) element of the $[i,j]$ block is $\mathbf{R}_{\boldsymbol{\vartheta}}[i,j]_{(j,i)} = b_{ij}$. All remaining elements are 0. By permuting rows and columns of $\mathbf{R}_{\boldsymbol{\vartheta}}$, it can be brought into block-diagonal form. Then it consists only of 1×1 blocks with elements d_i and 2×2 blocks $\begin{bmatrix} a_{ij} & b_{ij} \\ b_{ji} & a_{ji} \end{bmatrix}$. Hence, $\mathbf{R}_{\boldsymbol{\vartheta}}$ can be easily inverted resulting in a block-diagonal matrix where all 1×1 and 2×2 blocks are individuallay inverted as long as $d_i \neq 0$ and $a_{ij}a_{ji} - b_{ij}b_{ji} \neq 0$. The result is

$$\mathbf{R}_{\boldsymbol{\vartheta}}^{-1} = \sum_{i=1}^{N} \frac{1}{d_i} \mathbf{L}_{ii} \otimes \mathbf{L}_{ii} + \sum_{i=1}^{N} \sum_{\substack{j=1 \\ j \neq i}}^{N} \frac{a_{ji}}{a_{ij}a_{ji} - b_{ij}b_{ji}} \mathbf{L}_{ii} \otimes \mathbf{L}_{jj} + \sum_{i=1}^{N} \sum_{\substack{j=1 \\ j \neq i}}^{N} \frac{-b_{ij}}{a_{ij}a_{ji} - b_{ij}b_{ji}} \mathbf{L}_{ij} \otimes \mathbf{L}_{ji} \quad (27)$$

$$= \sum_{i=1}^{N} f_i \mathbf{L}_{ii} \otimes \mathbf{L}_{ii} + \sum_{i=1}^{N} \sum_{\substack{j=1 \\ j \neq i}}^{N} \left(g_{ij} \mathbf{L}_{ii} \otimes \mathbf{L}_{jj} + h_{ij} \mathbf{L}_{ij} \otimes \mathbf{L}_{ji}\right) \quad (28)$$

with

$$f_i = \frac{1}{d_i} = \frac{\eta_i - 1}{(\eta_i - 1)^2 - |\beta_i - 1|^2}, \quad (29)$$

$$g_{ij} = \frac{a_{ji}}{a_{ij}a_{ji} - b_{ij}b_{ji}} = \frac{\kappa_j(\kappa_i\kappa_j - 1) - |\gamma_i\xi_j|^2\kappa_i}{u_{ij}}, \quad (30)$$

$$h_{ij} = \frac{-b_{ij}}{a_{ij}a_{ji} - b_{ij}b_{ji}} = \frac{\gamma_j^*\xi_i^*\kappa_j + \gamma_i\xi_j\kappa_i}{u_{ij}}. \quad (31)$$

where $u_{ij} = (\kappa_i\kappa_j - 1)^2 + |\gamma_i\gamma_j\xi_i\xi_j - 1|^2 - 1 - \kappa_i^2|\gamma_i\xi_j|^2 - \kappa_j^2|\gamma_j\xi_i|^2$. This means that $\mathrm{var}(\hat{G}_{ii})$ and $\mathrm{var}(\hat{G}_{ij})$ of $\hat{\mathbf{G}} = \hat{\mathbf{W}}\mathbf{A}$ are lower bounded by the (i,i)-th and (j,j)-th element of the (i,i)-th block of $L^{-1}\mathbf{R}_{\boldsymbol{\vartheta}}^{-1}$:

$$\mathrm{var}(\hat{G}_{ii}) \geq \frac{1}{L} f_i, = \frac{1}{L} \frac{\eta_i - 1}{(\eta_i - 1)^2 - |\beta_i - 1|^2}, \quad (32)$$

$$\mathrm{var}(\hat{G}_{ij}) \geq \frac{1}{L} g_{ij} = \frac{1}{L} \frac{\kappa_j(\kappa_i\kappa_j - 1) - |\gamma_i\xi_j|^2\kappa_i}{u_{ij}}. \quad (33)$$

Note that $L^{-1}\mathbf{R}_{\boldsymbol{\vartheta}}^{-1}$ is the iCRB for $\boldsymbol{\vartheta}$ as in (11). In order to get the complete iCRB for $\begin{bmatrix} \boldsymbol{\vartheta} \\ \boldsymbol{\vartheta}^* \end{bmatrix}$ as in (10), we also calculate $\mathbf{P}_{\boldsymbol{\vartheta}} = -\mathbf{R}_{\boldsymbol{\vartheta}}^{-1}\mathbf{Q}_{\boldsymbol{\vartheta}} = -\mathbf{R}_{\boldsymbol{\vartheta}}^{-1}\mathbf{M}_2^*\mathbf{M}_1^{-1}$. It can be shown in a similar way

$$\mathbf{P}_{\boldsymbol{\vartheta}} = \sum_{i=1}^{N} \tilde{f}_i \mathbf{L}_{ii} \otimes \mathbf{L}_{ii} + \sum_{i=1}^{N} \sum_{\substack{j=1 \\ j \neq i}}^{N} \left(\tilde{g}_{ij} \mathbf{L}_{ii} \otimes \mathbf{L}_{jj} + \tilde{h}_{ij} \mathbf{L}_{ij} \otimes \mathbf{L}_{ji}\right) \quad (34)$$

with

$$\tilde{f}_i = -\frac{f_i(\beta_i - 1)^*}{\eta_i - 1} = -\frac{(\beta_i - 1)^*}{(\eta_i - 1)^2 - |\beta_i - 1|^2}, \quad (35)$$

$$\tilde{g}_{ij} = -\frac{g_{ij}\gamma_j^*\xi_i^* + h_{ij}}{\kappa_i} = -\frac{(\kappa_j^2 - |\gamma_i\xi_j|^2)\gamma_j^*\xi_i^*\gamma_i\xi_j}{u_{ij}}, \quad (36)$$

$$\tilde{h}_{ij} = -\frac{g_{ij} + \gamma_i^*\xi_j^* h_{ij}}{\kappa_j} = -\frac{\kappa_i\kappa_j - 1 + (\gamma_j\xi_i\gamma_i\xi_j)^*}{u_{ij}}. \quad (37)$$

Note that according to (28) and (34) the iCRB for $\mathbf{G} = \mathbf{W}\mathbf{A}$ has a nice decoupling property: the iCRB for G_{ii} only depends on the distribution of source i and the iCRB for G_{ij} only depends on the distribution of sources i and j and not on other sources. Note that (32) and (33) cannot be used as a bound for real ICA since the FIM would be singular.

B. CRB for the demixing matrix $\mathbf{W}$

Starting with the iCRB $L^{-1}\mathbf{R}_{\boldsymbol{\vartheta}}^{-1}$ for the stacked gain matrix $\boldsymbol{\vartheta} = \mathrm{vec}((\mathbf{W}\mathbf{A})^T) = (\mathbf{I} \otimes \mathbf{A}^T)\,\mathrm{vec}(\mathbf{W}^T)$, it is now straightforward to derive the CRB for the stacked demixing matrix $\boldsymbol{\theta} = \mathrm{vec}(\mathbf{W}^T) = (\mathbf{I} \otimes \mathbf{A}^T)^{-1}\boldsymbol{\vartheta} = (\mathbf{I} \otimes \mathbf{W}^T)\boldsymbol{\vartheta}$. Since $\boldsymbol{\theta}$ is a linear function of $\boldsymbol{\vartheta}$,

$$\mathrm{cov}(\hat{\boldsymbol{\theta}}) \geq L^{-1}\mathbf{R}_{\boldsymbol{\theta}}^{-1} = L^{-1}(\mathbf{I} \otimes \mathbf{W}^T)\mathbf{R}_{\boldsymbol{\vartheta}}^{-1}(\mathbf{I} \otimes \mathbf{W}^*) \quad (38)$$

holds for any unbiased estimator $\hat{\boldsymbol{\theta}}$ for $\boldsymbol{\theta}$.

Using the properties of the Kronecker product in (65), we obtain

$$\mathbf{R}_{\boldsymbol{\theta}}^{-1} = \sum_{i=1}^{N} f_i \mathbf{L}_{ii} \otimes \mathbf{W}^T\mathbf{L}_{ii}\mathbf{W}^* + \sum_{i=1}^{N} \sum_{\substack{j=1 \\ j \neq i}}^{N} g_{ij} \mathbf{L}_{ii} \otimes \mathbf{W}^T\mathbf{L}_{jj}\mathbf{W}^* + \sum_{i=1}^{N} \sum_{\substack{j=1 \\ j \neq i}}^{N} h_{ij} \mathbf{L}_{ij} \otimes \mathbf{W}^T\mathbf{L}_{ji}\mathbf{W}^* \quad (39)$$

	both circular $\gamma_i = \gamma_j = 0$	both noncircular, identical noncircularities $\gamma_i = \gamma_j \neq 0$	different noncircularities $\gamma_i \neq \gamma_j$
both Gaussian	Sec. IV-A	Sec. IV-B	Sec. IV-B
both non-Gaussian, identical non-Gaussianities	Sec. IV-A	Sec. V-A	Sec. V-B
different non-Gaussianities	Sec. IV-A	Sec. V-C	Sec. V-C

TABLE I: Different combinations of the distribution of two sources i and j

$\mathbf{R}_{\boldsymbol{\theta}}^{-1}$ is a block-matrix whose (i,j) block is given by

$$\mathbf{R}_{\boldsymbol{\theta}}^{-1}[i,j] = \begin{cases} f_i \mathbf{w}_i \mathbf{w}_i^H + \sum_{\substack{l=1 \\ l \neq i}}^N g_{il} \mathbf{w}_l \mathbf{w}_l^H & i = j \\ h_{ij} \mathbf{w}_j \mathbf{w}_i^H & i \neq j. \end{cases} \tag{40}$$

$\mathbf{P}_{\boldsymbol{\theta}} = -\mathbf{R}_{\boldsymbol{\theta}}^{-1}\mathbf{Q}_{\boldsymbol{\theta}}$ can be calculated in a similar way.

IV. Special cases of the iCRB

In the previous section, we derived the iCRB for the gain matrix $\mathbf{G} = \mathbf{WA}$ for the general complex case. Below, we study some special cases of the iCRB. For reasons of completeness, the iCRB for real ICA from [3], [10], [11] is summarized in Appendix F. Due to the decoupling property of the iCRB mentioned in the previous section, the iCRB for G_{ii} and G_{ij} depends only on the distribution of the sources i and j. Both sources can be Gaussian or non-Gaussian, with identical or different non-Gaussianities. Both sources can be circular complex or noncircular complex, with identical or different noncircularity indices. Table 1 summarizes these 9 cases and shows which of them are discussed in which section.

A. All sources are circular complex

If all sources are circular complex, $\gamma_i = 0$ and $\beta_i = \eta_i$ (see Lemma 3). Due to the phase ambiguity in circular complex ICA, the Fisher information for the diagonal elements G_{ii} is 0 and hence their iCRB does not exist. However, we can constrain G_{ii} to be real and derive the constrained CRB [30] (see also Appendix D) for G_{ii}: As noted at the end of Sec. III-A, G_{ii} is decoupled from G_{ij} and G_{jj} and hence it is sufficient to consider the constrained CRB for G_{ii} alone. Let $\theta = G_{ii}$. The constraint $\theta \in \mathbb{R}$ can be formulated as $f(\theta) = \theta - \theta^* = 0$. We then need to calculate the Jacobian matrix $\mathbf{F}(\theta) = \begin{bmatrix} \partial f/\partial\theta & \partial f/\partial\theta^* \\ \partial f^*/\partial\theta & \partial f^*/\partial\theta^* \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ -1 & 1 \end{bmatrix}$ and find an unit-length vector $\mathbf{U}$ in the null space of $\mathbf{F}(\theta)$, i.e. $\mathbf{FU} = \mathbf{0}$. We choose $\mathbf{U} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \end{bmatrix}^T$. The CRB for the constrained parameter $\theta = G_{ii} \in \mathbb{R}$ is then

$$\begin{bmatrix} \mathrm{var}(\theta) & \mathrm{pvar}(\theta) \\ \mathrm{pvar}^*(\theta) & \mathrm{var}(\theta) \end{bmatrix} \geq \frac{1}{L} \mathbf{U} \left(\mathbf{U}^H \begin{bmatrix} \mathcal{I}_\theta & \mathcal{P}_\theta \\ \mathcal{P}_\theta^* & \mathcal{I}_\theta \end{bmatrix} \mathbf{U} \right)^{-1} \mathbf{U}^H$$
$$= \frac{1}{4L(\eta_i - 1)} \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \tag{41}$$

where $\mathcal{I}_\theta = \eta_i - 1 = \mathcal{P}_\theta$ and $\eta_i = \mathrm{E}\left[|s_i|^2 |\varphi_i(s_i)|^2\right]$. This means

$$\mathrm{var}(\hat{G}_{ii}) \geq \frac{1}{4L(\eta_i - 1)}. \tag{42}$$

The bound in (42) is valid for a phase-constrained G_{ii} such that $G_{ii} \in \mathbb{R}$. Eq. (42) looks similar to the real case (97) except for a factor of 4 since η_i is defined using Wirtinger derivatives instead of real derivatives.

For $\mathrm{var}(\hat{G}_{ij})$ we get from (33)

$$\mathrm{var}(\hat{G}_{ij}) \geq \frac{1}{L} \frac{\kappa_j}{\kappa_i \kappa_j - 1}, \tag{43}$$

which again looks the same as in the real case (98). However, in the complex case, κ_i is defined using the Wirtinger derivative instead of real derivative. Furthermore, in the complex case κ measures the non-Gaussianity and noncircularity whereas in the real case κ measures only the non-Gaussianity.

If source i and j are both circular Gaussian, $\kappa_i = \kappa_j = 1$ and $\mathrm{var}(\hat{G}_{ij}) \to \infty$. This corresponds to the known fact that circular complex Gaussian sources cannot be separated by ICA.

B. All sources are noncircular complex Gaussian

If all sources are noncircular Gaussian with different $\gamma_i \in \mathbb{R}$, it can be shown using the expressions for κ, ξ, η and β from Appendix E, (93)-(96) with $c = 1$ that

$$\mathrm{var}(\hat{G}_{ii}) \geq \frac{1}{L} \frac{1}{4\gamma_i^2}, \tag{44}$$

$$\mathrm{var}(\hat{G}_{ij}) \geq \frac{1}{L} \frac{\gamma_i^2 + \gamma_j^2 - 2\gamma_i^2\gamma_j^2}{(\gamma_j^2 - \gamma_i^2)^2} (1 - \gamma_i^2)$$
$$= \frac{1 - \gamma_i^2}{2L} \left[\frac{1 - \gamma_i\gamma_j}{(\gamma_i - \gamma_j)^2} + \frac{1 + \gamma_i\gamma_j}{(\gamma_i + \gamma_j)^2} \right]. \tag{45}$$

Note that (45) is exactly the same result as obtained by Yeredor in [21] for the performance analysis of the SUT, i.e. our result shows that for noncircular Gaussian sources the SUT is indeed asymptotically optimal.

If all sources are noncircular Gaussian with identical γ_i, it can be shown that the iCRB for G_{ij} does not exist because $\gamma_j^2 - \gamma_i^2 \to 0$. This confirms the result obtained in [5], [19] which showed that ICA fails for two or more noncircular Gaussian signals with same γ_i.

V. Results for generalized Gaussian distribution

In order to verify the CRB derived in the previous sections, we now study complex ICA with noncircular complex generalized Gaussian distributed (GGD) sources. We choose this family of parametric pdf since it enables an analytical calculation of the CRB. The pdf of such a noncircular complex source s with zero mean, variance $\mathrm{E}[|s|^2] = 1$ and noncircularity index $\gamma \in [0, 1]$ can be written as [35]

$$p(s, s^*) = \frac{c\alpha \cdot \exp\left(-\left[\frac{\alpha/2}{\gamma^2 - 1} \left(\gamma s^2 + \gamma s^{*2} - 2ss^* \right) \right]^c \right)}{\pi\Gamma(1/c)(1 - \gamma^2)^{1/2}},$$

where $\alpha = \Gamma(2/c)/\Gamma(1/c)$ and $\Gamma(\cdot)$ is the Gamma function. The shape parameter $c > 0$ varies the form of the pdf from super-Gaussian ($c < 1$) to sub-Gaussian ($c > 1$). For $c = 1$, the pdf is Gaussian. $0 \leq \gamma \leq 1$ controls the noncircularity of the pdf. The four parameters κ, β, η, ξ required to calculate the CRB are derived in Appendix E.

For the simulation study, we consider $N = 3$ sources with random mixing matrices $\mathbf{A}$. The real and imaginary part of all elements of $\mathbf{A}$ are independent and uniformly distributed in $[-1, 1]$. We conducted 100 experiments with different random matrices $\mathbf{A}$ and considered two different ICA estimators[2]: Complex ML-ICA [19] and complex ICA by entropy bound minimization (ICA-EBM) [20]. Complex ML-ICA finds the demixing matrix $\mathbf{W}$ by maximizing the likelihood $p(\mathbf{x}; \mathbf{W})$ and hence requires knowledge of the pdf $p_i(s_i)$ of each source i. In the simulations, ML-ICA uses for each source i a separate, fixed noncircular complex GGD with the known shape parameter c_i and noncircularity index γ_i. Optimization is performed using natural-gradient ascent [36] with a normalized step-size for each source. Complex ICA-EBM, on the other hand, is based on a flexible entropy estimator which can adapt to a wide range of distributions and hence can be employed without prior knowledge of the source pdfs. [20] provides a performance comparison of eight different complex ICA algorithms in terms of percentage of failures, average interference-to-signal ratio and average CPU time. It was shown that ICA-EBM provides a good tradeoff between separation performance and computational complexity. We additionally consider complex ML-ICA, since per definition it reaches the CRB asymptotically.

In this paper, we want to compare the separation performance of ICA with respect to the iCRB and hence we define the performance metric as in [10]: After running an ICA algorithm, we correct the permutation ambiguity of the estimated demixing matrix and calculate the signal-to-interference ratio (SIR) averaged over all N sources:

$$\mathrm{SIR} = \frac{1}{N}\sum_{i=1}^{N}\frac{\mathrm{E}\left[|G_{ii}|^2\right]}{\sum_{j\neq i}\mathrm{E}[|G_{ij}|^2]} = \frac{1}{N}\sum_{i=1}^{N}\frac{1+\mathrm{var}(G_{ii})}{\sum_{j\neq i}\mathrm{var}(G_{ij})}. \quad (46)$$

In this definition, the averaging over simulation trials takes place before taking the ratio.

In practice, the accuracy of the estimated demixing matrix depends not only on the optimization *cost function* but also on the optimization *algorithm* used to implement the estimator: In some rare cases, complex ML-ICA based on natural-gradient ascent converges to a local maximum of the likelihood and yields a lower SIR value than ICA-EBM. To overcome this problem, we initialized ML-ICA from the solution obtained by ICA-EBM which is close to the optimal solution. MATLAB code for the simulations is available at http://www.iss.uni-stuttgart.de/institut/mitarbeiter_alt/loesch/complex_ICA_CRB/index.html.

A quite general optimization algorithm for learning in Lie groups was recently introduced in [37]. Furthermore, a way to average several ICA solutions in order to modify the variance has been described in [38]. The different solutions can either be obtained by running the same algorithm more than once or by running different algorithms on the same dataset. However, algorithmic improvements are beyond the scope of this paper.

[2]Note that many alternative ICA estimators such as [7], [12]–[18] exist.

A. All sources are identically distributed

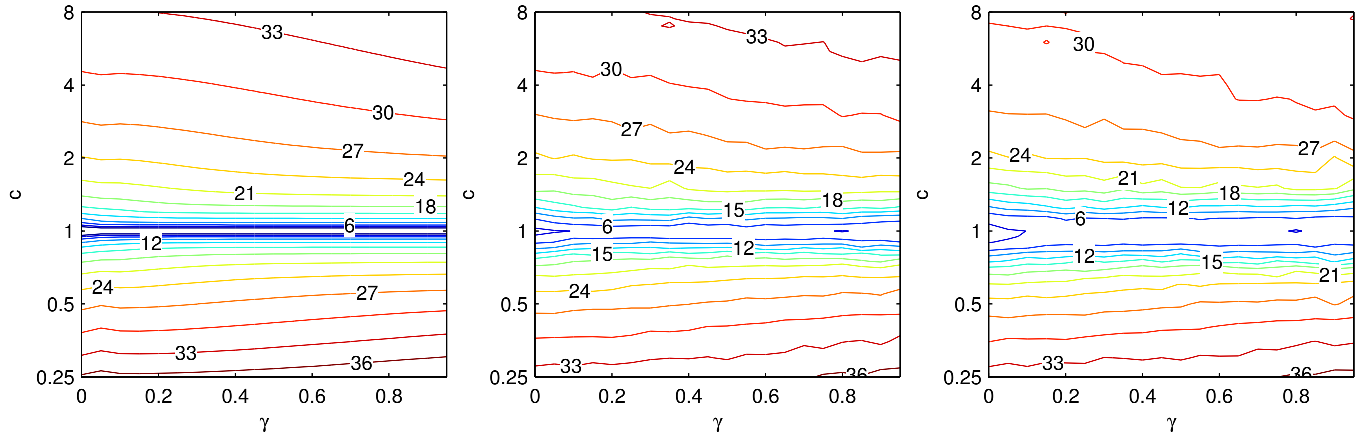
First, we study the performance when all sources are identically distributed with the same shape parameter c and the same noncircularity index γ. Fig. 2 shows the results: The SIR given by the iCRB increases with increasing non-Gaussianity ($c \to \infty$ or $c \to 0$). For $c \approx 1$, SIR is low since (nearly) Gaussian sources with the same noncircularity index γ cannot be separated by ICA. For $c \neq 1$, the SIR also increases with increasing noncircularity γ, but much slower since all sources have the same noncircularity γ. Clearly, both ICA algorithms work quite well except for $c \approx 1$ (Gaussian). Their SIR comes quite close to the CRB and ML-ICA slightly outperforms ICA-EBM, especially for strongly sub-Gaussian sources ($c > 1$). This is due to the fact that ML-ICA uses nonlinearities matched to the source distributions while ICA-EBM uses a linear combination of prespecified nonlinear functions. Note that ICA-EBM allows one to select nonlinearities for approximating the source entropy. Hence if prior knowledge about the source distributions is available, it can be incorporated into ICA-EBM thus improving its performance.

B. All sources have the same shape parameter but different noncircularities

Now we study the performance when all sources follow a GGD with the same shape parameter $c_i = c$ but have different noncircularity indices $\gamma_i = (i-1)\Delta\gamma$. Fig. 3 shows that the SIR given by the iCRB increases both with increasing non-Gaussianity as well as increasing difference in noncircularity indices $\Delta\gamma$ even for $c = 1$. This is due to the fact that Gaussian sources with distinct noncircularity indices γ can be separated by ICA. In the previous subsection, all sources had the same noncircularity γ and hence a separation was not possible for $c = 1$. Fig. 3 shows that ML-ICA again outperforms ICA-EBM which is again due to the fact that ML-ICA uses for each source s_i a nonlinearity $\varphi_i(s_i)$ matched to its pdf $p_i(s_i)$. Since ML-ICA and ICA-EBM use both non-Gaussianity and noncircularity for separation, the contour lines in Fig. 3 (b) and Fig. 3 (c) stay close to those of the iCRB in Fig. 3 (a).

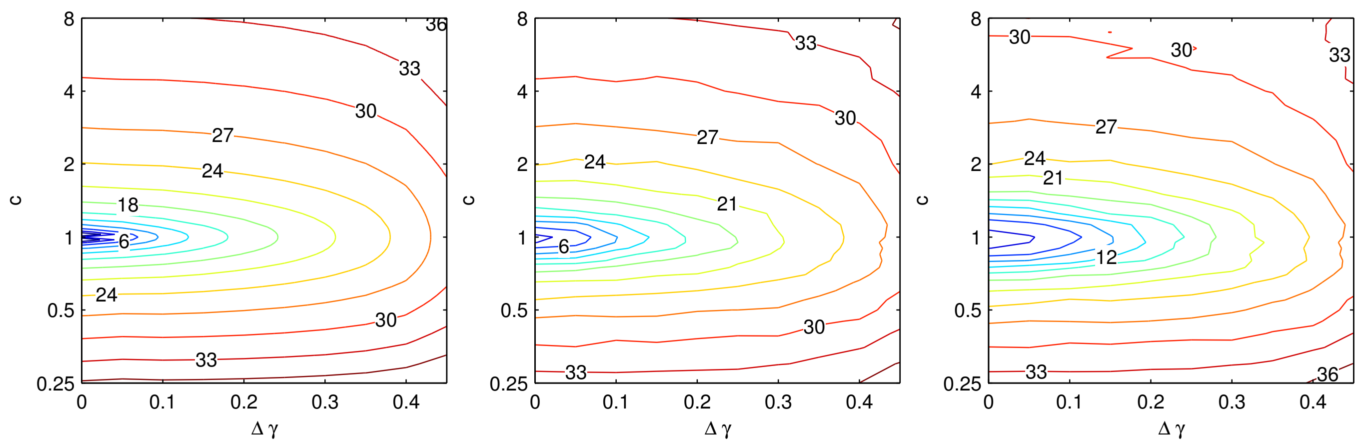
C. All sources have different shape parameters

Now we study the performance when the sources follow a GGD with different shape parameters $c_1 = 1, c_2 = c, c_3 = 1/c$. Fig. 4 shows the result when all sources have the same noncircularity indices $\gamma_i = \gamma$ while in Fig. 5 sources have different noncircularity indices $\gamma_i = (i-1)\Delta\gamma$. In both cases ML-ICA and ICA-EBM come quite close to the CRB but ML-ICA slightly outperforms ICA-EBM. The reason is again that ML-ICA uses for each source s_i a nonlinearity $\varphi_i(s_i)$ matched to its pdf $p_i(s_i)$ whereas the nonlinearities used in ICA-EBM are fixed a priori. When sources differ in shape parameter *and* noncircularity (see Fig. 5) the SIR increases both with increasing non-Gaussianity of source 2 and 3 (i.e. $c < 1$) as well as increasing difference in noncircularity indices $\Delta\gamma$.



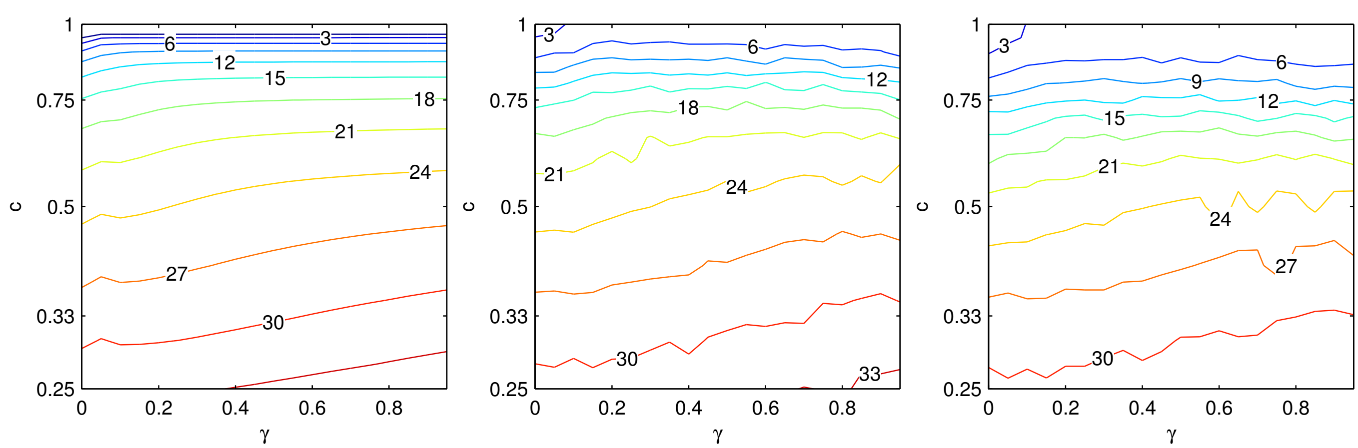
(a) iCRB (b) ML-ICA (c) ICA-EBM

Fig. 2: Comparison of signal-to-interference ratio [dB] of different ICA estimators with CRB, sample size $L = 1000$, all sources follow a generalized Gaussian distribution with $c_i = c$ and $\gamma_i = \gamma$



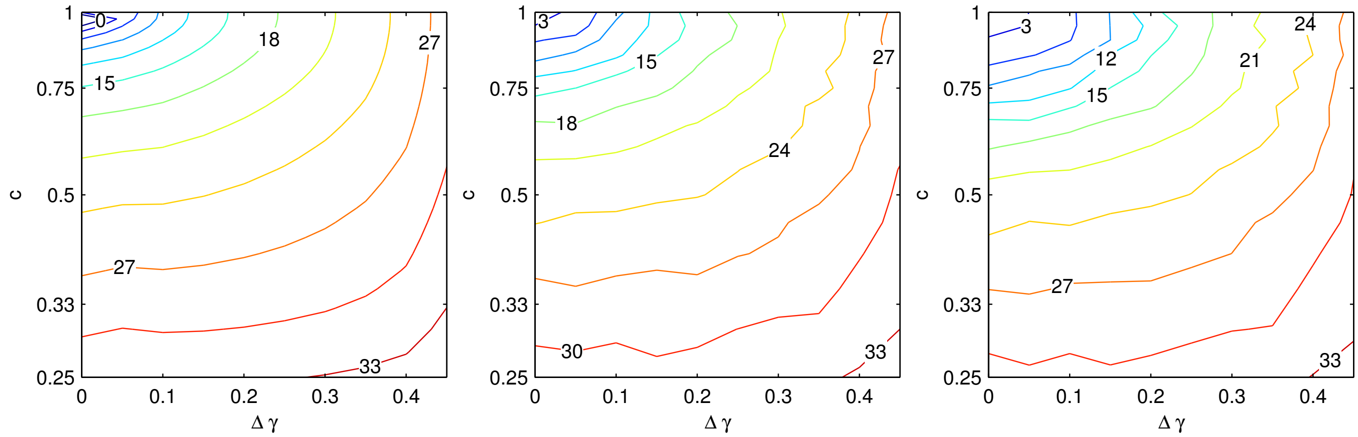
(a) iCRB (b) ML-ICA (c) ICA-EBM

Fig. 3: Comparison of signal-to-interference ratio [dB] of different ICA estimators with CRB, sample size $L = 1000$, all sources follow a generalized Gaussian distribution with $c_i = c$, $\gamma_i = (i-1)\Delta\gamma$



(a) iCRB (b) ML-ICA (c) ICA-EBM

Fig. 4: Comparison of signal-to-interference ratio [dB] of different ICA estimators with iCRB, sample size $L = 1000$, all sources follow a generalized Gaussian distribution with $c_1 = 1, c_2 = c, c_3 = 1/c$, $\gamma_i = \gamma$

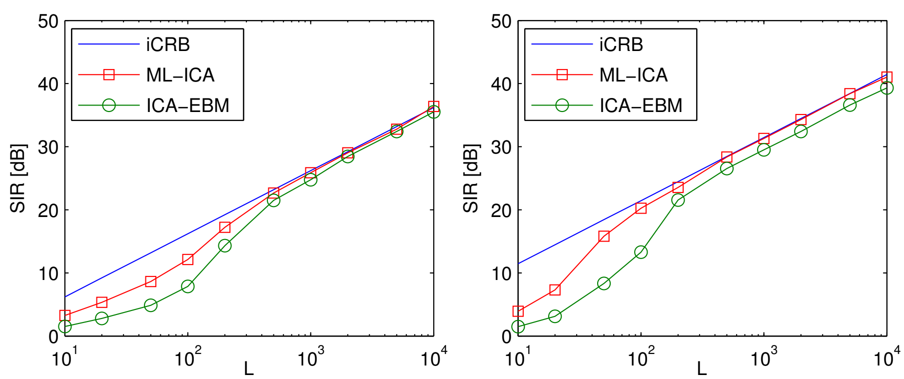


(a) iCRB

(b) ML-ICA

(c) ICA-EBM

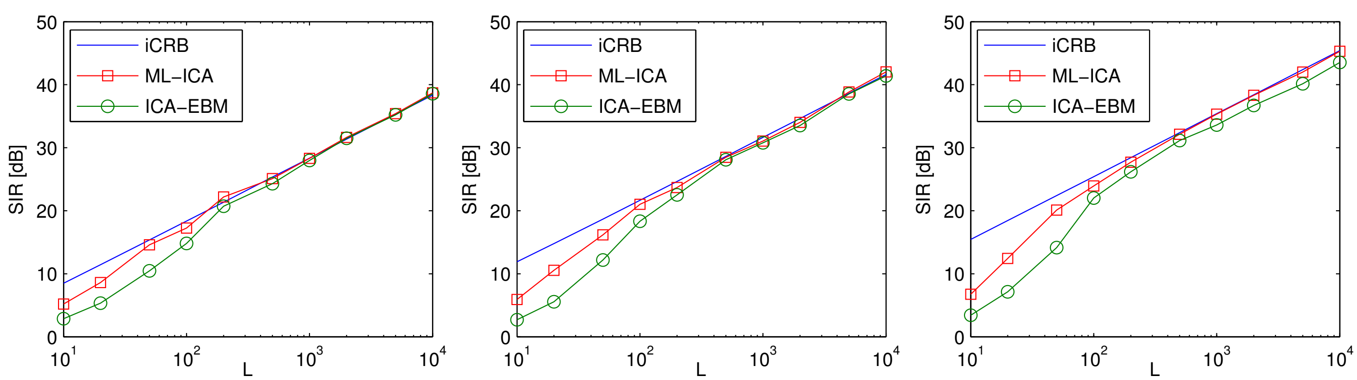
Fig. 5: Comparison of signal-to-interference ratio [dB] of different ICA estimators with iCRB, sample size $L = 1000$, all sources follow a generalized Gaussian distribution with $c_1 = 1, c_2 = c, c_3 = 1/c$, $\gamma_i = (i-1)\Delta\gamma$



(a) $c = 0.5$

(b) $c = 6$

Fig. 6: Performance as a function of sample size L, circular GGD sources



(a) $c = 1$, $\Delta\gamma = 0.45$

(b) $c = 0.5$, $\Delta\gamma = 0.45$

(c) $c = 6$, $\Delta\gamma = 0.45$

Fig. 7: Performance as a function of sample size L, noncircular GGD sources with $c_i = c$ and $\gamma_i = (i-1)\Delta\gamma$

D. Performance as a function of the sample size

Here we study the performance of ICA algorithms as a function of the sample size L. Fig. 6 shows that for circular non-Gaussian sources and a limited sample size $L < 1000$, ML-ICA outperforms ICA-EBM. Fig. 7(a)-(c) shows that also for noncircular signals with distinct noncircularity indices and small sample size, ML-ICA outperforms ICA-EBM. However, for a large sample size ($L \geq 1000$), ICA-EBM comes close to the performance of ML-ICA except for strongly sub-Gaussian sources (see Fig. 6 (b), Fig. 7 (c) and also Sec. V-A).

VI. Extension of iCRB to nonwhite or nonstationary sources

Here, we extend the iCRB to the case where source signals are not temporally iid. Our derivation follows the derivation for the real case by Cardoso in chapter 4 of [3]. The real iCRB for general temporally non-iid sources as well as Gaussian temporally non-iid sources is briefly summarized in Appendix F. We collect L temporally non-iid samples of the source signals in an $N \times L$ matrix $\mathbf{S}$. The mixing model (1) becomes $\mathbf{X} = \mathbf{AS}$, where $\mathbf{X}$ is an $N \times L$ matrix. Let $\mathbf{s}_i$ denote the i-th row vector of $\mathbf{S}$ and $s_i(l)$ the (i, l)-th element of $\mathbf{S}$. Since different source signals (rows of $\mathbf{S}$) are independent, the pdf of $\mathbf{S}$ is given as $p_{\mathbf{S}}(\mathbf{S}) = \prod_{i=1}^{N} p_{\mathbf{s}_i}(\mathbf{s}_i)$. The pdf of $\mathbf{X}$ is then given by $p(\mathbf{X}; \boldsymbol{\theta}) = |\det(\mathbf{W})|^{2L} \prod_{i=1}^{N} p_{\mathbf{s}_i}(\mathbf{w}_i^T \mathbf{X})$. By using matrix derivatives [22], [24], [31], we obtain

$$\begin{aligned}\frac{\partial}{\partial \mathbf{W}^H} \ln p(\mathbf{X}; \boldsymbol{\theta}) &= L\mathbf{A}^* - \mathbf{X}^* \boldsymbol{\psi}^T(\mathbf{WX}) \\ &= L\mathbf{A}^* \left(\mathbf{I} - \frac{1}{L} \mathbf{S} \boldsymbol{\psi}^H(\mathbf{S}) \right)^* \end{aligned} \tag{47}$$

where the i-th row of $\boldsymbol{\psi}$ depends only on $\mathbf{s}_i$ and is given by $[\boldsymbol{\psi}]_i = \boldsymbol{\varphi}_i(\mathbf{s}_i)$. The l-th element of the row vector $\boldsymbol{\varphi}_i(\mathbf{s}_i)$ is given by

$$[\boldsymbol{\varphi}_i(\mathbf{s}_i)]_l = -\frac{\partial}{\partial s_i^*(l)} \ln p_{\mathbf{s}_i}(\mathbf{s}_i). \tag{48}$$

To derive the iCRB for $\mathbf{G} = \mathbf{WA}$, we calculate $\mathbf{M}_1$ and $\mathbf{M}_2$ in a similar way as in the iid case:

$$\begin{aligned}\mathbf{M}_1 &= L^2 \operatorname{E}\left[\operatorname{vec}\left\{ \mathbf{I} - \frac{1}{L} \mathbf{S} \boldsymbol{\psi}^H(\mathbf{S}) \right\} \operatorname{vec}\left\{ \mathbf{I} - \frac{1}{L} \mathbf{S} \boldsymbol{\psi}^H(\mathbf{S}) \right\}^H \right], \\ \mathbf{M}_2 &= L^2 \operatorname{E}\left[\operatorname{vec}\left\{ \mathbf{I} - \frac{1}{L} \mathbf{S} \boldsymbol{\psi}^H(\mathbf{S}) \right\} \operatorname{vec}\left\{ \mathbf{I} - \frac{1}{L} \mathbf{S} \boldsymbol{\psi}^H(\mathbf{S}) \right\}^T \right].\end{aligned} \tag{49}$$

A convenient notation is also $\mathbf{M}_1 = \operatorname{cov}(L \operatorname{vec}\{\mathbf{I} - \frac{1}{L} \mathbf{S} \boldsymbol{\psi}^H(\mathbf{S})\})$ and $\mathbf{M}_2 = \operatorname{pcov}(L \operatorname{vec}\{\mathbf{I} - \frac{1}{L} \mathbf{S} \boldsymbol{\psi}^H(\mathbf{S})\})$.

Due to the source independence, $\mathbf{M}_1$ and $\mathbf{M}_2$ have a sparse decoupled structure as in the iid case: They consist of N 1×1 blocks important for the iCRB of G_{ii} and $\frac{N(N-1)}{2}$ 2×2 blocks important for the iCRB of G_{ij}. All the remaining elements are zero. We define

$$\begin{aligned}\eta_i - 1 &= L \operatorname{var}\left(1 - \frac{1}{L} \mathbf{s}_i \boldsymbol{\varphi}_i^H(\mathbf{s}_i) \right) = L \operatorname{E}\left[\left| \frac{1}{L} \mathbf{s}_i \boldsymbol{\varphi}_i^H(\mathbf{s}_i) \right|^2 - 1 \right], \\ \beta_i - 1 &= L \operatorname{pvar}\left(1 - \frac{1}{L} \mathbf{s}_i \boldsymbol{\varphi}_i^H(\mathbf{s}_i) \right) = L \operatorname{E}\left[\left(\frac{1}{L} \mathbf{s}_i \boldsymbol{\varphi}_i^H(\mathbf{s}_i) \right)^2 - 1 \right], \\ \rho_{ij} &= \frac{1}{L} \operatorname{E}\left[\left| \mathbf{s}_j \boldsymbol{\varphi}_i^H(\mathbf{s}_i) \right|^2 \right] = \frac{1}{L} \operatorname{E} \operatorname{tr}\left(\mathbf{s}_j \boldsymbol{\varphi}_i^H(\mathbf{s}_i) \boldsymbol{\varphi}_i(\mathbf{s}_i) \mathbf{s}_j^H \right) \\ &= \frac{1}{L} \operatorname{tr}\left(\operatorname{E}\left[\mathbf{s}_j^H \mathbf{s}_j \right] \operatorname{E}\left[\boldsymbol{\varphi}_i^H(\mathbf{s}_i) \boldsymbol{\varphi}_i(\mathbf{s}_i) \right] \right) = \frac{1}{L} \operatorname{tr}\left(\mathbf{R}_j \boldsymbol{\kappa}_i \right), \\ \chi_{ij} &= \frac{1}{L} \operatorname{E}\left[\left(\mathbf{s}_j \boldsymbol{\varphi}_i^H(\mathbf{s}_i) \right)^2 \right] = \frac{1}{L} \operatorname{E} \operatorname{tr}\left(\mathbf{s}_j \boldsymbol{\varphi}_i^H(\mathbf{s}_i) \boldsymbol{\varphi}_i^*(\mathbf{s}_i) \mathbf{s}_j^T \right), \\ &= \frac{1}{L} \operatorname{tr}\left(\operatorname{E}\left[\mathbf{s}_j^T \mathbf{s}_j \right] \operatorname{E}\left[\boldsymbol{\varphi}_i^H(\mathbf{s}_i) \boldsymbol{\varphi}_i^*(\mathbf{s}_i) \right] \right) = \frac{1}{L} \operatorname{tr}\left(\mathbf{P}_j \boldsymbol{\xi}_i \right),\end{aligned} \tag{50}$$

with

$$\begin{aligned}&\mathbf{R}_j = \operatorname{E}\left[\mathbf{s}_j^H \mathbf{s}_j \right], \quad \mathbf{P}_j = \operatorname{E}\left[\mathbf{s}_j^T \mathbf{s}_j \right], \\ &\boldsymbol{\kappa}_i = \operatorname{E}\left[\boldsymbol{\varphi}_i^H(\mathbf{s}_i) \boldsymbol{\varphi}_i(\mathbf{s}_i) \right], \quad \boldsymbol{\xi}_i = \operatorname{E}\left[\boldsymbol{\varphi}_i^H(\mathbf{s}_i) \boldsymbol{\varphi}_i^*(\mathbf{s}_i) \right].\end{aligned} \tag{51}$$

Furthermore, it holds

$$\begin{aligned}\frac{1}{L} \operatorname{E}\left[\mathbf{s}_j \boldsymbol{\varphi}_i^H(\mathbf{s}_i) \left(\mathbf{s}_i \boldsymbol{\varphi}_j^H(\mathbf{s}_j) \right)^* \right] &= \frac{1}{L} \operatorname{E} \operatorname{tr}\left(\mathbf{s}_j \boldsymbol{\varphi}_i^H(\mathbf{s}_i) \mathbf{s}_i^* \boldsymbol{\varphi}_j^T(\mathbf{s}_j) \right) \\ &= \frac{1}{L} \operatorname{tr}\left(\operatorname{E}\left[\boldsymbol{\varphi}_j^T(\mathbf{s}_j) \mathbf{s}_j \right] \operatorname{E}\left[\boldsymbol{\varphi}_i^H(\mathbf{s}_i) \mathbf{s}_i^* \right] \right) = 0, \\ \frac{1}{L} \operatorname{E}\left[\mathbf{s}_j \boldsymbol{\varphi}_i^H(\mathbf{s}_i) \left(\mathbf{s}_i \boldsymbol{\varphi}_j^H(\mathbf{s}_j) \right) \right] &= \frac{1}{L} \operatorname{E} \operatorname{tr}\left(\mathbf{s}_j \boldsymbol{\varphi}_i^H(\mathbf{s}_i) \mathbf{s}_i \boldsymbol{\varphi}_j^H(\mathbf{s}_j) \right) \\ &= \frac{1}{L} \operatorname{tr}\left(\operatorname{E}\left[\boldsymbol{\varphi}_j^H(\mathbf{s}_j) \mathbf{s}_j \right] \operatorname{E}\left[\boldsymbol{\varphi}_i^H(\mathbf{s}_i) \mathbf{s}_i \right] \right) = 1.\end{aligned} \tag{52}$$

Hence, it can be shown that $\mathbf{M}_1$ and $\mathbf{M}_2$ are given by

$$\begin{aligned}\mathbf{M}_1 &= L \left(\sum_{i=1}^{N} (\eta_i - 1) \mathbf{L}_{ii} \otimes \mathbf{L}_{ii} + \sum_{i=1}^{N} \sum_{\substack{j=1 \\ j \neq i}}^{N} \rho_{ij} \mathbf{L}_{ii} \otimes \mathbf{L}_{jj} \right), \\ \mathbf{M}_2 &= L \left(\sum_{i=1}^{N} (\beta_i - 1) \mathbf{L}_{ii} \otimes \mathbf{L}_{ii} + \sum_{i=1}^{N} \sum_{\substack{j=1 \\ j \neq i}}^{N} \chi_{ij} \mathbf{L}_{ii} \otimes \mathbf{L}_{jj} \right. \\ &\qquad \left. + \sum_{i=1}^{N} \sum_{\substack{j=1 \\ j \neq i}}^{N} \mathbf{L}_{ij} \otimes \mathbf{L}_{ji} \right).\end{aligned} \tag{53}$$

The remaining steps in the derivation of the iCRB for the non-iid case are the same as in the iid case except for scaling of the FIM by the number of samples L: In the iid case $\mathbf{M}_1$ and $\mathbf{M}_2$ in (82) and (87) were defined for a single sample and hence need to be scaled by L, whereas in the non-iid case the definition of $\mathbf{M}_1$ and $\mathbf{M}_2$ in (53) already contains this factor. Hence, we can directly use the results for the iid case from (32) and (33). We only have to replace κ_i by ρ_{ij} and $\gamma_j \xi_i$ by χ_{ij} to obtain

$$\operatorname{var}(\hat{G}_{ii}) \geq \frac{1}{L} \frac{\eta_i - 1}{(\eta_i - 1)^2 - |\beta_i - 1|^2}, \tag{54}$$

$$\operatorname{var}(\hat{G}_{ij}) \geq \frac{1}{L} \frac{\rho_{ji}(\rho_{ij}\rho_{ji} - 1) - |\chi_{ji}|^2 \rho_{ij}}{u_{ij}}. \tag{55}$$

where $u_{ij} = (\rho_{ij}\rho_{ji} - 1)^2 + |\chi_{ij}\chi_{ji} - 1|^2 - 1 - \rho_{ij}^2 |\chi_{ji}|^2 - \rho_{ji}^2 |\chi_{ij}|^2$ and η_i, β_i, ρ_{ij} and χ_{ij} are defined in (50).

A. Temporally non-iid Gaussian sources

The pdf of a multivariate complex Gaussian *column* random vector $\mathbf{s}$ can be written as [39]

$$p(\mathbf{s}) = \frac{\exp\left(-\mathbf{s}^H\mathbf{C}^{-1}\mathbf{s} + \Re\left\{\mathbf{s}_i^T\mathbf{R}^{-*}\mathbf{P}^*\mathbf{C}^{-1}\mathbf{s}\right\}\right)}{\pi^L\left[\det(\mathbf{R}\cdot\mathbf{C}^*)\right]^{1/2}} \quad (56)$$

with $\mathbf{R} = \mathrm{E}[\mathbf{s}\mathbf{s}^H]$, $\mathbf{P} = \mathrm{E}[\mathbf{s}\mathbf{s}^T]$ and $\mathbf{C} = \mathbf{R} - \mathbf{P}\mathbf{R}^{-*}\mathbf{P}^*$. Correspondingly, the pdf $p(\mathbf{s}_i)$ for the *row* vector $\mathbf{s}_i$ containing L temporally non-iid samples of the noncircular complex Gaussian source i can be written as

$$p(\mathbf{s}_i) = \frac{\exp\left(-\mathbf{s}_i\mathbf{C}_i^{-1}\mathbf{s}_i^H + \Re\left\{\mathbf{s}_i\mathbf{R}_i^{-1}\mathbf{P}_i^*\mathbf{C}_i^{-*}\mathbf{s}_i^T\right\}\right)}{\pi^L\left[\det(\mathbf{R}_i\cdot\mathbf{C}_i^*)\right]^{1/2}} \quad (57)$$

where now $\mathbf{R}_i = \mathrm{E}[\mathbf{s}_i^H\mathbf{s}_i]$, $\mathbf{P}_i = \mathrm{E}[\mathbf{s}_i^T\mathbf{s}_i]$ and $\mathbf{C}_i = \mathbf{R}_i - \mathbf{P}_i^*\mathbf{R}_i^{-*}\mathbf{P}_i$.

It can be shown that $\boldsymbol{\varphi}_i(\mathbf{s}_i)$ is given by

$$\boldsymbol{\varphi}_i(\mathbf{s}_i) = (\mathbf{s}_i - \mathbf{s}_i^*\mathbf{R}_i^{-*}\mathbf{P}_i)\mathbf{C}^{-1}. \quad (58)$$

Using a multivariate version of Lemma 1 from Appendix B we obtain from (50)

$$\begin{aligned}
\eta_i - 1 &= \frac{1}{L}\operatorname{tr}\left(\mathrm{E}\left[\mathbf{s}_i^H\mathbf{s}_i\boldsymbol{\varphi}_i^H(\mathbf{s}_i)\boldsymbol{\varphi}_i(\mathbf{s}_i)\right]\right) - L \\
&= \frac{1}{L}\operatorname{tr}\left(\mathrm{E}\left[\frac{\partial\,\mathbf{s}_i^H\mathbf{s}_i\boldsymbol{\varphi}_i^H(\mathbf{s}_i)}{\partial\mathbf{s}_i^*}\right]\right) - L = \frac{1}{L}\operatorname{tr}\left(\mathbf{R}_i\mathbf{C}_i^{-1}\right), \\
\beta_i - 1 &= \frac{1}{L}\operatorname{tr}\left(\mathrm{E}\left[\mathbf{s}_i^T\mathbf{s}_i\boldsymbol{\varphi}_i^H(\mathbf{s}_i)\boldsymbol{\varphi}_i^*(\mathbf{s}_i)\right]\right) - L \\
&= \frac{1}{L}\operatorname{tr}\left(\mathrm{E}\left[\frac{\partial\,\mathbf{s}_i^T\mathbf{s}_i\boldsymbol{\varphi}_i^H(\mathbf{s}_i)}{\partial\mathbf{s}_i}\right]\right) - L \\
&= 1 - \frac{1}{L}\operatorname{tr}\left(\mathbf{P}_i\mathbf{C}_i^{-1}\mathbf{P}_i^*\mathbf{R}_i^{-*}\right). \quad (59)
\end{aligned}$$

For $\operatorname{var}(\hat{G}_{ij})$ we calculate using (58)

$$\boldsymbol{\kappa}_i = \mathrm{E}\left[\boldsymbol{\varphi}_i^H(\mathbf{s}_i)\boldsymbol{\varphi}_i(\mathbf{s}_i)\right] = \mathbf{C}_i^{-H} = \mathbf{C}_i^{-1}, \quad (60)$$

$$\boldsymbol{\xi}_i = \mathrm{E}\left[\boldsymbol{\varphi}_i^H(\mathbf{s}_i)\boldsymbol{\varphi}_i^*(\mathbf{s}_i)\right] = -\mathbf{C}_i^{-1}\mathbf{P}_i^*\mathbf{R}_i^{-*} \quad (61)$$

and

$$\rho_{ij} = \frac{1}{L}\operatorname{tr}(\mathbf{R}_j\boldsymbol{\kappa}_i) = \frac{1}{L}\operatorname{tr}(\mathbf{R}_j\mathbf{C}_i^{-1}), \quad (62)$$

$$\chi_{ij} = \frac{1}{L}\operatorname{tr}(\mathbf{P}_j\boldsymbol{\xi}_i) = -\frac{1}{L}\operatorname{tr}(\mathbf{P}_j\mathbf{C}_i^{-1}\mathbf{P}_i^*\mathbf{R}_i^{-*}). \quad (63)$$

It can be shown that the FIM for G_{ij} becomes singular, i.e. $\operatorname{var}(\hat{G}_{ij}) \to \infty$ in (55) if $\mathbf{R}_j = \alpha\mathbf{R}_i$ and $\mathbf{P}_j = \pm\alpha\mathbf{P}_i$ with the same $\alpha > 0$. If we assume that all sources s_i are rotated such that $\mathrm{E}[s_i^2(1)] \geq 0$, this reduces to $\mathbf{R}_j = \alpha\mathbf{R}_i$ and $\mathbf{P}_j = \alpha\mathbf{P}_i$ with the same $\alpha > 0$. Compared to the case of real temporally non-iid Gaussian sources [3], [34] where the FIM becomes singular if $\mathbf{R}_j = \mathrm{E}[\mathbf{s}_j^T\mathbf{s}_j] = \alpha\mathbf{R}_i = \alpha\,\mathrm{E}[\mathbf{s}_i^T\mathbf{s}_i]$, we obtain a more complicated condition in the noncircular complex case.

For the special case of circular complex sources, $\mathbf{P}_i = \mathbf{P}_j = \mathbf{0}$ and $\mathbf{C}_i = \mathbf{R}_i$. Hence $\eta_i = \beta_i = 2$, $\chi_{ij} = \chi_{ji} = 0$ and $\rho_{ij} = L^{-1}\operatorname{tr}(\mathbf{R}_j\mathbf{R}_i^{-1})$. As a consequence, the Fisher information for G_{ii} becomes singular and $\operatorname{var}(\hat{G}_{ii}) \to \infty$ in (54) as in the temporally iid case. Then we could obtain a constrained CRB as in Sec. IV-A. Furthermore, (55) can be simplified to $\operatorname{var}(G_{ij}) \geq L^{-1}\rho_{ji}\cdot(\rho_{ij}\rho_{ji} - 1)^{-1}$. The FIM for G_{ij} becomes singular, i.e. $\operatorname{var}(G_{ij}) \to \infty$ if $\mathbf{R}_j = \alpha\mathbf{R}_i$ which is similar to the real case except that $\mathbf{R}_i$ and $\mathbf{R}_j$ are defined with $(\cdot)^H$ instead of $(\cdot)^T$.

A different special case is stationarity of the sources for which $\mathbf{R}_i$ and $\mathbf{P}_i$ are Toeplitz matrices completely characterized by the correlation function $r_i(k) = \mathrm{E}[s_i(l+k)s_i^*(l)]$ and pseudo-correlation function $\gamma_i(k) = \mathrm{E}[s_i(l+k)s_i(l)]$ of each source i with $k \in [0, L-1]$. Under the above rotation convention, the FIM for G_{ij} becomes singular if, for two source i and j, $r_j(k) = \alpha r_i(k)$ and $\gamma_j(k) = \alpha\gamma_i(k)$ $\forall k \in [0, L-1]$ with the same $\alpha > 0$. [40] proposed a joint diagonalization approach for the separation of noncircular complex stationary sources and showed that the demixing matrix is identifiable, i.e. sources are separable as long as the vectors $\mathbf{r}_i = [r_i(0), \cdots, r_i(L-1), \gamma_i(0), \cdots, \gamma_i(L-1)]^T$ are distinct for all sources i. If we reverse this condition, i.e. $\mathbf{r}_j = \alpha\mathbf{r}_i$ for some source $j \neq i$ and $\alpha > 0$, the FIM for G_{ij} becomes singular and source i and j are nonseparable. Note that $\mathbf{r}_j = \alpha\mathbf{r}_i$ is equivalent to our condition $r_j(k) = \alpha r_i(k)$ and $\gamma_j(k) = \alpha\gamma_i(k)$ $\forall k \in [0, L-1]$.

Finally, for temporally uncorrelated but nonstationary sources, $\mathbf{R}_i = \operatorname{diag}\left(\sigma_i^2(1), \cdots, \sigma_i^2(L)\right)$ and $\mathbf{P}_i = \operatorname{diag}\left(\gamma_i(1), \cdots, \gamma_i(L)\right)$ where now $\sigma_i^2(l)$ and $\gamma_i(l)$ denote the time-variant variance and pseudo-variance of each source i. The FIM then becomes singular if $\sigma_j^2(l) = \alpha\sigma_i^2(l)$ and $\gamma_j(l) = \alpha\gamma_i(l)$ $\forall l \in [1, L]$ with the same $\alpha > 0$.

VII. Conclusion

In this paper, we have derived the CRB for the noncircular complex ICA problem with temporally iid sources. The induced CRB (iCRB) for the gain matrix, i.e. the demixing-mixing-matrix product, depends on the distribution of the sources through five parameters, which can be easily calculated. The derived bound is valid for the general noncircular complex case and contains the circular complex and the noncircular complex Gaussian case as two special cases. The iCRB reflects the phase ambiguity in circular complex ICA. In that case, we derived a constrained CRB for a phase-constrained demixing matrix. Simulation results using two ICA algorithms have shown that for sources following a noncircular complex generalized Gaussian distribution, these algorithms can achieve a signal-to-interference ratio (SIR) close to that of the iCRB. The complex ML-ICA algorithm, which uses for each source a nonlinearity matched to its pdf, outperforms ICA-EBM especially for small sample sizes. However, for ML-ICA the pdfs of the sources must be known whereas no such knowledge is required for ICA-EBM. Hence, for practical applications where the pdfs of the sources might be unknown ICA-EBM is an adequate algorithm whose performance comes quite close to the iCRB for large enough sample size L. Finally, we have also shown how to extend the iCRB from the temporally iid case to the general non-iid case, considered the special case of complex non-iid Gaussian sources and discussed the differences to the real case.

Appendix A
Useful matrix algebra

As in [11], we make use of some matrix algebra in the derivation of the CRB. Let $\mathbf{L}_{ij} = \mathbf{e}_i\mathbf{e}_j^T$ denote an $N \times N$

matrix with a 1 at the (i,j) position and 0's elsewhere. $\mathbf{e}_i$ is a length-N vector with a 1 at the i-th position and 0's elsewhere. It is useful to note that

$$\mathbf{A}\mathbf{L}_{ij}\mathbf{A}^T = \mathbf{a}_i\mathbf{a}_j^T, \quad \mathbf{L}_{ij}\mathbf{L}_{kl} = \mathbf{0} \text{ for } j \neq k, \quad \mathbf{L}_{ij}\mathbf{L}_{jl} = \mathbf{L}_{il}. \tag{64}$$

A useful rule for the Kronecker product $\otimes$ is

$$(\mathbf{A}\otimes\mathbf{B})(\mathbf{C}\otimes\mathbf{D}) = (\mathbf{AC})\otimes(\mathbf{BD}), \tag{65}$$

where all matrices are such that the operation is properly defined. We also note that any $N^2 \times N^2$ block matrix $\mathbf{A}$ containing $N\times N$ blocks $\mathbf{A}[i,j]$ can be written using its $N\times N$ diagonal blocks $\mathbf{A}[i,i]$ and $N\times N$ off-diagonal blocks $\mathbf{A}[i,j], i\neq j$ as follows:

$$\mathbf{A} = \sum_{i=1}^{N} \mathbf{L}_{ii}\otimes\mathbf{A}[i,i] + \sum_{i=1}^{N}\sum_{j=1,\ j\neq i}^{N} \mathbf{L}_{ij}\otimes\mathbf{A}[i,j]. \tag{66}$$

By combining the properties (64) and (65), we obtain (67).

Appendix B
Some additional lemmas used in the derivation of the CRB

For the Lemmas and Corollaries in this section, we use the definition of $\varphi(s)$ from (14) and the definitions of η,κ,β and ξ from (22)-(25). The following Lemma provides a very useful property of the function $\varphi(s)$.

Lemma 1. *For any real differentiable function $g(s)$, it holds $\mathrm{E}\left[g(s)\varphi^*(s)\right] = \mathrm{E}\left[\frac{\partial g}{\partial s}\right]$ and $\mathrm{E}\left[g(s)\varphi(s)\right] = \mathrm{E}\left[\frac{\partial g}{\partial s^*}\right]$ as long as $g(s)p(s)\to 0$ for $s_R\to\pm\infty$ or $s_I\to\pm\infty$.*

Proof: The first part of the lemma, $\mathrm{E}\left[g(s)\varphi^*(s)\right] = \mathrm{E}\left[\frac{\partial g}{\partial s}\right]$, has been proven in [22]. Since $\varphi(s) = -\frac{\partial \ln p(s)}{\partial s^*}$ and $\varphi^*(s) = \left(-\frac{\partial \ln p(s)}{\partial s^*}\right)^* = -\frac{\partial \ln p(s)}{\partial s}$, the second part of the lemma, $\mathrm{E}\left[g(s)\varphi(s)\right] = \mathrm{E}\left[\frac{\partial g}{\partial s^*}\right]$, follows immediately. ■

Since sources are independent and zero mean, we get the identities in Corrolaries 1-3 from Lemma 1:

Corollary 1. *It holds*

$$\mathrm{E}[s_i\varphi_j^*(s_j)] = \begin{cases} 1 & i=j \\ 0 & i\neq j \end{cases}, \tag{68}$$

$$\mathrm{E}[s_i\varphi_j(s_j)] = 0. \tag{69}$$

The next corollary provides alternative forms of the parameters κ,ξ,η,β which are sometimes easier to calculate than the ones from (22)-(25):

Corollary 2. *It holds*

$$\kappa = \mathrm{E}\left[|\varphi(s)|^2\right] = \mathrm{E}\left[\frac{\partial\varphi}{\partial s}\right] = \mathrm{E}\left[\frac{\partial\varphi^*}{\partial s^*}\right], \tag{70}$$

$$\xi = \mathrm{E}\left[(\varphi^*(s))^2\right] = \mathrm{E}\left[\frac{\partial\varphi^*}{\partial s}\right], \tag{71}$$

$$\eta = \mathrm{E}\left[|s|^2|\varphi(s)|^2\right] = \mathrm{E}\left[|s|^2\frac{\partial\varphi}{\partial s}\right] + 1 = \mathrm{E}\left[|s|^2\frac{\partial\varphi^*}{\partial s^*}\right] + 1 \tag{72}$$

$$\beta = \mathrm{E}\left[s^2(\varphi^*(s))^2\right] = \mathrm{E}\left[s^2\frac{\partial\varphi^*}{\partial s}\right] + 2. \tag{73}$$

Corollary 3. *Using Corollary 1, $\eta>1$, since*

$$\begin{aligned}\eta &= \mathrm{E}\left[|s\varphi^*(s)|^2\right] = |\mathrm{E}\left[s\varphi^*(s)\right]|^2 + \mathrm{var}\left(s\varphi^*(s)\right) \\ &= 1 + \mathrm{var}\left(s\varphi^*(s)\right) > 1.\end{aligned} \tag{74}$$

Now we provide some lemmas for distributions satisfying certain symmetry properties.

Lemma 2. *Both ξ_i and β_i are real if $p(-s_R,s_I)=p(s_R,s_I)$ or $p(s_R,-s_I)=p(s_R,s_I)$, i.e. $p(s_R,s_I)$ is symmetric in s_R or s_I.*

Proof:

$$\begin{aligned}\xi &= \mathrm{E}\left[(\varphi^*(s))^2\right] \\ &= \int\int \frac{1}{4}\left(\frac{\partial p(s_R,s_I)}{\partial s_R} - j\frac{\partial p(s_R,s_I)}{\partial s_I}\right)^2 \frac{1}{p(s_R,s_I)} ds_R ds_I \\ &= \frac{1}{4}\int\int \left[\left(\frac{\partial p(s_R,s_I)}{\partial s_R}\right)^2 - \left(\frac{\partial p(s_R,s_I)}{\partial s_I}\right)^2 \right. \\ &\qquad \left. -2j\frac{\partial p(s_R,s_I)}{\partial s_R}\frac{\partial p(s_R,s_I)}{\partial s_I}\right]\frac{1}{p(s_R,s_I)} ds_R ds_I.\end{aligned} \tag{75}$$

If $p(s_R,s_I)$ is symmetric in s_R or s_I, i.e. $p(-s_R,s_I) = p(s_R,s_I)$ or $p(s_R,-s_I) = p(s_R,s_I)$, then $h(s_R,s_I) = \frac{\partial p}{\partial s_R}\frac{\partial p}{\partial s_I}$ is antisymmetric in s_R or s_I, i.e. $h(-s_R,s_I) = -h(s_R,s_I)$ or $h(s_R,-s_I) = -h(s_R,s_I)$. Hence $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(s_R,s_I)\frac{1}{p(s_R,s_I)} ds_R ds_I = 0$ and $\xi\in\mathbb{R}$. A similar proof can be made for β. ■

Lemma 3. *If s_i is circular complex, $\beta_i=\eta_i$.*

Proof: For circular $s = s_R + js_I$, $p(s_R,s_I) = g(s_R^2+s_I^2)$. Let $f(r^2) = f(s_R^2+s_I^2) = \ln p(s_R,s_I)$. It holds

$$\begin{aligned}\eta &= \frac{1}{4}\,\mathrm{E}\left[(s_R^2+s_I^2)\left(\left(\frac{\partial f}{\partial s_R}\right)^2 + \left(\frac{\partial f}{\partial s_I}\right)^2\right)\right], \\ \beta &= \frac{1}{4}\,\mathrm{E}\left[(s_R^2 - s_I^2 + 2js_Rs_I)\left(\left(\frac{\partial f}{\partial s_R}\right)^2 - \left(\frac{\partial f}{\partial s_I}\right)^2 \right.\right. \\ &\qquad \left.\left. -2j\left(\frac{\partial f}{\partial s_R}\right)\left(\frac{\partial f}{\partial s_I}\right)\right)\right], \\ &= \frac{1}{4}\,\mathrm{E}\left[(s_R^2 - s_I^2)\left(\left(\frac{\partial f}{\partial s_R}\right)^2 - \left(\frac{\partial f}{\partial s_I}\right)^2\right) \right. \\ &\qquad \left. +4s_Rs_I\left(\frac{\partial f}{\partial s_R}\right)\left(\frac{\partial f}{\partial s_I}\right)\right], \\ 4(\eta-\beta) &= 2\,\mathrm{E}\left[s_R^2\left(\frac{\partial f}{\partial s_I}\right)^2 + s_I^2\left(\frac{\partial f}{\partial s_R}\right)^2 \right. \\ &\qquad \left. -2s_Rs_I\left(\frac{\partial f}{\partial s_R}\right)\left(\frac{\partial f}{\partial s_I}\right)\right] = 0,\end{aligned}$$

where we used $\mathrm{E}\left[s_Rs_I\left(\left(\frac{\partial f}{\partial s_R}\right)^2 - \left(\frac{\partial f}{\partial s_I}\right)^2\right)\right] = 0$ and

$$\left(\sum_{i=1}^{N}\mathbf{L}_{ii}\otimes\mathbf{L}_{ii}\right)\left(\sum_{i=1}^{N}\sum_{\substack{j=1\\j\neq i}}^{N}\mathbf{L}_{ii}\otimes\mathbf{L}_{jj}\right)=\mathbf{0},\quad\left(\sum_{i=1}^{N}\sum_{\substack{j=1\\j\neq i}}^{N}\mathbf{L}_{ii}\otimes\mathbf{L}_{jj}\right)\left(\sum_{i=1}^{N}\mathbf{L}_{ii}\otimes\mathbf{L}_{ii}\right)=\mathbf{0},$$

$$\left(\sum_{i=1}^{N}\mathbf{L}_{ii}\otimes\mathbf{L}_{ii}\right)\left(\sum_{i=1}^{N}\sum_{\substack{j=1\\j\neq i}}^{N}\mathbf{L}_{ij}\otimes\mathbf{L}_{ji}\right)=\mathbf{0},\quad\left(\sum_{i=1}^{N}\sum_{\substack{j=1\\j\neq i}}^{N}\mathbf{L}_{ij}\otimes\mathbf{L}_{ji}\right)\left(\sum_{i=1}^{N}\mathbf{L}_{ii}\otimes\mathbf{L}_{ii}\right)=\mathbf{0},$$

$$\left(\sum_{i=1}^{N}\sum_{\substack{j=1\\j\neq i}}^{N}\mathbf{L}_{ij}\otimes\mathbf{L}_{ji}\right)\left(\sum_{i=1}^{N}\sum_{\substack{j=1\\j\neq i}}^{N}c_{ij}\mathbf{L}_{ii}\otimes\mathbf{L}_{jj}\right)=\sum_{i=1}^{N}\sum_{\substack{j=1\\j\neq i}}^{N}c_{ji}\mathbf{L}_{ij}\otimes\mathbf{L}_{ji},$$

$$\left(\sum_{i=1}^{N}\sum_{\substack{j=1\\j\neq i}}^{N}c_{ij}\mathbf{L}_{ii}\otimes\mathbf{L}_{jj}\right)\left(\sum_{i=1}^{N}\sum_{\substack{j=1\\j\neq i}}^{N}\mathbf{L}_{ij}\otimes\mathbf{L}_{ji}\right)=\sum_{i=1}^{N}\sum_{\substack{j=1\\j\neq i}}^{N}c_{ij}\mathbf{L}_{ij}\otimes\mathbf{L}_{ji},$$

$$\left(\sum_{i=1}^{N}\sum_{\substack{j=1\\j\neq i}}^{N}c_{ij}\mathbf{L}_{ij}\otimes\mathbf{L}_{ji}\right)\left(\sum_{i=1}^{N}\sum_{\substack{j=1\\j\neq i}}^{N}\mathbf{L}_{ij}\otimes\mathbf{L}_{ji}\right)=\sum_{i=1}^{N}\sum_{\substack{j=1\\j\neq i}}^{N}c_{ij}\mathbf{L}_{ii}\otimes\mathbf{L}_{jj}. \tag{67}$$

$\mathrm{E}\Big[\left(s_R^2-s_I^2\right)\left(\frac{\partial f}{\partial s_R}\right)\left(\frac{\partial f}{\partial s_I}\right)\Big]=0$ in the third line and $\frac{\partial f}{\partial s_R}=2s_R\frac{\partial f(r^2)}{\partial r^2}$ and $\frac{\partial f}{\partial s_I}=2s_I\frac{\partial f(r^2)}{\partial r^2}$ in the last line. ■

APPENDIX C
DERIVATION OF THE iCRB FOR THE GAIN MATRIX $\mathbf{G}$

The derivation of the iCRB for the matrix $\mathbf{G}=\mathbf{WA}$ proceeds in three steps:

1) Calculate the matrices $\mathbf{M}_1$ and $\mathbf{M}_2$ in (17)
2) Calculate the matrix $\mathbf{R}_{\boldsymbol{\vartheta}}=(\mathbf{M}_1-\mathbf{M}_2\mathbf{M}_1^{-*}\mathbf{M}_2^*)^*$ in (20)
3) Invert $\mathbf{R}_{\boldsymbol{\vartheta}}$

Using $\mathrm{E}[\mathbf{s}\boldsymbol{\varphi}^H(\mathbf{s})]=\mathbf{I}$ (see Lemma 1), we can simplify $\mathbf{M}_1$ as

$$\begin{aligned}\mathbf{M}_1&=\mathrm{E}\left[\mathrm{vec}\{\mathbf{I}-\mathbf{s}\boldsymbol{\varphi}^H(\mathbf{s})\}\,\mathrm{vec}\{\mathbf{I}-\mathbf{s}\boldsymbol{\varphi}^H(\mathbf{s})\}^H\right]\\&=\mathrm{E}\left[\mathrm{vec}\{\mathbf{s}\boldsymbol{\varphi}^H(\mathbf{s})\}\,\mathrm{vec}\{\mathbf{s}\boldsymbol{\varphi}^H(\mathbf{s})\}^H\right]-\mathrm{vec}\{\mathbf{I}\}\,\mathrm{vec}\{\mathbf{I}\}^H\\&=\boldsymbol{\Omega}_1-\mathrm{vec}\{\mathbf{I}\}\,\mathrm{vec}\{\mathbf{I}\}^H.\end{aligned} \tag{76}$$

$\boldsymbol{\Omega}_1=\mathrm{E}\left[\mathrm{vec}\{\mathbf{s}\boldsymbol{\varphi}^H(\mathbf{s})\}\,\mathrm{vec}\{\mathbf{s}\boldsymbol{\varphi}^H(\mathbf{s})\}^H\right]$ is an $N^2\times N^2$ block matrix. The (i,i) block $\boldsymbol{\Omega}_1[i,i]=\mathrm{E}\left[\mathbf{s}\mathbf{s}^H|\varphi_i(s_i)|^2\right]$ is diagonal since the components of $\mathbf{s}$ are independent and zero mean. The diagonal elements $\boldsymbol{\Omega}_1[i,i]_{(j,j)}$ are given by

$$\boldsymbol{\Omega}_1[i,i]_{(j,j)}=\begin{cases}\mathrm{E}\left[|s_i|^2|\varphi_i(s_i)|^2\right]=\eta_i & i=j\\ \mathrm{E}\left[|s_j|^2|\varphi_i(s_i)|^2\right]=\mathrm{E}\left[|\varphi_i(s_i)|^2\right]=\kappa_i & i\neq j\end{cases} \tag{77}$$

according to (22) and (23) and due to the independence of s_j and s_i. κ_i and η_i are real. The (i,j) block $\boldsymbol{\Omega}_1[i,j]$ $(i\neq j)$ can be calculated as $\boldsymbol{\Omega}_1[i,j]=\mathrm{E}\left[\mathbf{s}\mathbf{s}^H\varphi_i^*(s_i)\varphi_j(s_j)\right]$. It has 1 at entry (i,j) and 0 at entry (j,i), since

$$\begin{aligned}\boldsymbol{\Omega}_1[i,j]_{(i,j)}&=\mathrm{E}\left[s_is_j^*\varphi_i^*(s_i)\varphi_j(s_j)\right]\\&=\mathrm{E}\left[s_i\varphi_i^*(s_i)\right]\mathrm{E}\left[s_j^*\varphi_j(s_j)\right]=1,\end{aligned} \tag{78}$$

$$\begin{aligned}\boldsymbol{\Omega}_1[i,j]_{(j,i)}&=\mathrm{E}\left[s_js_i^*\varphi_j(s_j)\varphi_i^*(s_i)\right]\\&=\mathrm{E}\left[s_i^*\varphi_i^*(s_i)\right]\mathrm{E}\left[s_j\varphi_j(s_j)\right]=0,\end{aligned} \tag{79}$$

due to Corollary 1. All other entries of $\boldsymbol{\Omega}_1[i,j]$ are zero since the components of $\mathbf{s}$ are independent and zero mean. Using Appendix A, we can write $\boldsymbol{\Omega}_1[i,i]=\eta_i\mathbf{L}_{ii}+\kappa_i\sum_{j\neq i}\mathbf{L}_{jj}$ and $\boldsymbol{\Omega}_1[i,j]=\mathbf{L}_{ij}$. Hence $\boldsymbol{\Omega}_1$ can be written as

$$\begin{aligned}\boldsymbol{\Omega}_1=\sum_{i=1}^{N}\eta_i\mathbf{L}_{ii}\otimes\mathbf{L}_{ii}&+\sum_{i=1}^{N}\sum_{\substack{j=1\\j\neq i}}^{N}\kappa_i\mathbf{L}_{ii}\otimes\mathbf{L}_{jj}\\&+\sum_{i=1}^{N}\sum_{\substack{j=1\\j\neq i}}^{N}\mathbf{L}_{ij}\otimes\mathbf{L}_{ij}.\end{aligned} \tag{80}$$

Using

$$\mathrm{vec}\{\mathbf{I}\}\,\mathrm{vec}\{\mathbf{I}\}^H=\sum_{i=1}^{N}\mathbf{L}_{ii}\otimes\mathbf{L}_{ii}+\sum_{i=1}^{N}\sum_{\substack{j=1\\j\neq i}}^{N}\mathbf{L}_{ij}\otimes\mathbf{L}_{ij}, \tag{81}$$

we can simplify $\mathbf{M}_1$ as

$$\mathbf{M}_1=\sum_{i=1}^{N}(\eta_i-1)\mathbf{L}_{ii}\otimes\mathbf{L}_{ii}+\sum_{i=1}^{N}\sum_{\substack{j=1\\j\neq i}}^{N}\kappa_i\mathbf{L}_{ii}\otimes\mathbf{L}_{jj}. \tag{82}$$

We note that $\mathbf{M}_1$ is a real diagonal matrix.

$\mathbf{M}_2$ can be calculated similarly:

$$\begin{aligned}\mathbf{M}_2&=\mathrm{E}\left[\mathrm{vec}\{\mathbf{I}-\mathbf{s}\boldsymbol{\varphi}^H(\mathbf{s})\}\,\mathrm{vec}\{\mathbf{I}-\mathbf{s}\boldsymbol{\varphi}^H(\mathbf{s})\}^T\right]\\&=\mathrm{E}\left[\mathrm{vec}\{\mathbf{s}\boldsymbol{\varphi}^H(\mathbf{s})\}\,\mathrm{vec}\{\mathbf{s}\boldsymbol{\varphi}^H(\mathbf{s})\}^T\right]-\mathrm{vec}\{\mathbf{I}\}\,\mathrm{vec}\{\mathbf{I}\}^T\\&=\boldsymbol{\Omega}_2-\mathrm{vec}\{\mathbf{I}\}\,\mathrm{vec}\{\mathbf{I}\}^T.\end{aligned} \tag{83}$$

$\boldsymbol{\Omega}_2=\mathrm{E}\left[\mathrm{vec}\{\mathbf{s}\boldsymbol{\varphi}^H(\mathbf{s})\}\,\mathrm{vec}\{\mathbf{s}\boldsymbol{\varphi}^H(\mathbf{s})\}^T\right]$ is an $N^2\times N^2$ block matrix. The (i,i) block $\boldsymbol{\Omega}_2[i,i]=\mathrm{E}\left[\mathbf{s}\mathbf{s}^T(\varphi_i^*(s_i))^2\right]$ is diagonal since the components of $\mathbf{s}$ are independent and zero mean. The diagonal elements $\mathbf{\Omega}_2[i,i]_{(j,j)}$ are given by

$$\mathbf{\Omega}_2[i,i]_{(j,j)} = \begin{cases} \mathrm{E}\left[s_i^2(\varphi_i^*(s_i))^2\right] = \beta_i & i=j \\ \mathrm{E}\left[s_j^2(\varphi_i^*(s_i))^2\right] = \mathrm{E}\left[s_j^2\right]\mathrm{E}\left[(\varphi_i^*(s_i))^2\right] & \\ \qquad = \gamma_j \xi_i & i \neq j \end{cases} \tag{84}$$

according to (24)-(26) and due to the independence of s_j and s_i. If $p(-s_R, s_I) = p(s_R, s_I)$ and $p(s_R, -s_I) = p(s_R, s_I)$, ξ_i and β_i are real (see Lemma 2). The (i,j) block $\mathbf{\Omega}_2[i,j]$ $(i \neq j)$ can be calculated as $\mathbf{\Omega}_2[i,j] = \mathrm{E}\left[\mathbf{s}\mathbf{s}^T \varphi_i^*(s_i)\varphi_j^*(s_j)\right]$. It has 1 at entry (i,j) and (j,i) since

$$\mathbf{\Omega}_2[i,j]_{(i,j)} = \mathbf{\Omega}_2[i,j]_{(j,i)} = \mathrm{E}\left[s_i\varphi_i^*(s_i)\right]\mathrm{E}\left[s_j^*\varphi_j(s_j)\right] = 1 \tag{85}$$

due to Corollary 1. All other entries of $\mathbf{\Omega}_2[i,j]$ are zero since the components of $\mathbf{s}$ are independent and zero mean. Using Appendix A, we can write $\mathbf{\Omega}_2[i,i] = \beta_i \mathbf{L}_{ii} + \sum_{j\neq i}\gamma_j\xi_i\mathbf{L}_{jj}$ and $\mathbf{\Omega}_2[i,j] = \mathbf{L}_{ij} + \mathbf{L}_{ji}$. Hence $\mathbf{\Omega}_2$ can be written as

$$\mathbf{\Omega}_2 = \sum_{i=1}^{N}\beta_i\mathbf{L}_{ii}\otimes\mathbf{L}_{ii} + \sum_{i=1}^{N}\sum_{\substack{j=1\\ j\neq i}}^{N}\gamma_j\xi_i\mathbf{L}_{ii}\otimes\mathbf{L}_{jj} + \sum_{i=1}^{N}\sum_{\substack{j=1\\ j\neq i}}^{N}(\mathbf{L}_{ij}\otimes\mathbf{L}_{ij} + \mathbf{L}_{ij}\otimes\mathbf{L}_{ji}). \tag{86}$$

Using $\mathrm{vec}\{\mathbf{I}\}\,\mathrm{vec}\{\mathbf{I}\}^T$ from (81), we can simplify $\mathbf{M}_2$ as

$$\mathbf{M}_2 = \sum_{i=1}^{N}(\beta_i - 1)\mathbf{L}_{ii}\otimes\mathbf{L}_{ii} + \sum_{i=1}^{N}\sum_{\substack{j=1\\ j\neq i}}^{N}\gamma_j\xi_i\mathbf{L}_{ii}\otimes\mathbf{L}_{jj} + \sum_{i=1}^{N}\sum_{\substack{j=1\\ j\neq i}}^{N}\mathbf{L}_{ij}\otimes\mathbf{L}_{ji}. \tag{87}$$

Since $\mathbf{M}_1$ is a real diagonal matrix, we can calculate $\mathbf{R}_{\boldsymbol{\vartheta}}$ from (20) as in (88).

After some simple but tedious calculations, we get using the properties from (67)

$$\mathbf{R}_{\boldsymbol{\vartheta}} = \sum_{i=1}^{N}d_i\mathbf{L}_{ii}\otimes\mathbf{L}_{ii} + \sum_{i=1}^{N}\sum_{\substack{j=1\\ j\neq i}}^{N}a_{ij}\mathbf{L}_{ii}\otimes\mathbf{L}_{jj} + \sum_{i=1}^{N}\sum_{\substack{j=1\\ j\neq i}}^{N}b_{ij}\mathbf{L}_{ij}\otimes\mathbf{L}_{ji} \tag{89}$$

where $d_i = \frac{(\eta_i-1)^2-|\beta_i-1|^2}{\eta_i-1}$, $a_{ij} = \kappa_i - \frac{|\gamma_j\xi_i|^2}{\kappa_i} - \frac{1}{\kappa_j}$ and $b_{ij} = -\left(\frac{\gamma_j^*\xi_i^*}{\kappa_i} + \frac{\gamma_i\xi_j}{\kappa_j}\right)$.

Appendix D
Complex Constrained CRB

In general, the constrained CRB for an N-dimensional complex parameter vector $\boldsymbol{\theta}$ under K equality constraints $\mathbf{0} = f(\boldsymbol{\theta}) \in \mathbb{C}^K$ can be derived by using the following steps [30]:

- Calculate the $2K\times 2N$ Jacobian matrix $\mathbf{F}(\boldsymbol{\theta})$ of $\mathbf{f}(\boldsymbol{\theta})$ with respect to $\boldsymbol{\theta}$ as $\mathbf{F}(\boldsymbol{\theta}) = \begin{bmatrix}\partial\mathbf{f}/\partial\boldsymbol{\theta} & \partial\mathbf{f}/\partial\boldsymbol{\theta}^* \\ \partial\mathbf{f}^*/\partial\boldsymbol{\theta} & \partial\mathbf{f}^*/\partial\boldsymbol{\theta}^*\end{bmatrix}$, where $\partial\mathbf{f}/\partial\boldsymbol{\theta} = \left[\frac{\partial f_i}{\partial\theta_j}\right]_{1\leq i\leq K, 1\leq j\leq N}$ and $\partial f_i/\partial\theta_j$ is defined using the Wirtinger derivatives in Sec. II.
- Find a $2N\times(2N-K)$ matrix $\mathbf{U}$ with orthonormal columns which span the null space of $\mathbf{F}$, i.e. $\mathbf{F}\mathbf{U} = \mathbf{0}$.
- The constrained CRB is then given as
$$\begin{bmatrix}\mathrm{cov}(\boldsymbol{\theta}) & \mathrm{pcov}(\boldsymbol{\theta}) \\ \mathrm{pcov}^*(\boldsymbol{\theta}) & \mathrm{cov}^*(\boldsymbol{\theta})\end{bmatrix} \geq \frac{1}{L}\mathbf{U}\left(\mathbf{U}^H\mathcal{J}_{\boldsymbol{\theta}}\mathbf{U}\right)^{-1}\mathbf{U}^H \tag{90}$$
where $(L^{-1}\mathcal{J}_{\boldsymbol{\theta}})^{-1}$ is the unconstrained CRB for estimating $\boldsymbol{\theta}$ from L iid samples.

Intuitively, the transform $\mathbf{U}^H\mathcal{J}_{\boldsymbol{\theta}}\mathbf{U}$ in (90) projects the Fisher information matrix $\mathcal{J}_{\boldsymbol{\theta}} = \begin{bmatrix}\mathcal{I}_{\boldsymbol{\theta}} & \mathcal{P}_{\boldsymbol{\theta}} \\ \mathcal{P}_{\boldsymbol{\theta}}^* & \mathcal{I}_{\boldsymbol{\theta}}^*\end{bmatrix}$ onto the subspace allowed by the constraints $f(\boldsymbol{\theta}) = \mathbf{0}$. The result of the inverse matrix $\left(\mathbf{U}^H\mathcal{J}_{\boldsymbol{\theta}}\mathbf{U}\right)^{-1}$ is then converted back to the original unconstrained space by the transformation $\mathbf{U}\cdot\left(\mathbf{U}^H\mathcal{J}_{\boldsymbol{\theta}}\mathbf{U}\right)^{-1}\cdot\mathbf{U}^H$.

Appendix E
Values of κ, ξ, β, η for complex GGD

The pdf of a noncircular complex GGD with zero mean, variance $\mathrm{E}[|s|^2] = 1$ and noncircularity index $\gamma \in [0,1]$ is given by

$$p(s,s^*) = \frac{c\alpha\cdot\exp\left(-\left[\frac{\alpha/2}{\gamma^2-1}\left(\gamma s^2 + \gamma s^{*2} - 2ss^*\right)\right]^c\right)}{\pi\Gamma(1/c)(1-\gamma^2)^{1/2}}, \tag{91}$$

where $\alpha = \Gamma(2/c)/\Gamma(1/c)$ and $\Gamma(\cdot)$ is the Gamma function. $\varphi(s,s^*) = -\frac{\partial}{\partial s^*}\ln p(s,s^*)$ is then given by

$$\varphi(s,s^*) = \frac{2c(\alpha/2)^c}{(\gamma^2-1)^c}\left(\gamma s^2 + \gamma(s^*)^2 - 2ss^*\right)^{c-1}(\gamma s^* - s). \tag{92}$$

By integration in polar coordinates, it can be shown that κ, ξ, β and η are given by

$$\kappa = \mathrm{E}\left[|\varphi(s)|^2\right] = \frac{c^2\Gamma(2/c)}{(1-\gamma^2)\Gamma^2(1/c)}, \tag{93}$$

$$\xi = \mathrm{E}\left[(\varphi^*(s))^2\right] = -\frac{c^2\gamma\Gamma(2/c)}{(1-\gamma^2)\Gamma^2(1/c)} = -\gamma\kappa, \tag{94}$$

$$\eta = \mathrm{E}\left[|s|^2|\varphi(s)|^2\right] = \frac{(c+1)\cdot(2-\gamma^2)}{2(1-\gamma^2)}, \tag{95}$$

$$\beta = \mathrm{E}\left[s^2(\varphi^*(s))^2\right] = \frac{(c+1)\cdot(2-3\gamma^2)}{2(1-\gamma^2)}. \tag{96}$$

Appendix F
Induced CRB for real ICA

Here, we briefly review the iCRB for real ICA [3], [10], [11], [34]. In the following, all real quantities q are denoted as $\bar{q}$. In the derivation of the iCRB for the real case and temporally iid sources $\bar{\varphi}(\bar{s}) = -\partial\ln p(\bar{s})/\partial\bar{s}$ and the parameters $\bar{\kappa} = E[\bar{\varphi}^2(\bar{s})]$, $\bar{\eta} = \mathrm{E}[\bar{s}^2\bar{\varphi}^2(\bar{s})] = 2 + \mathrm{E}\left[\bar{s}^2\frac{\partial\bar{\varphi}(\bar{s})}{\partial\bar{s}}\right]$ are defined using real derivatives. In [10], [11] it was shown that

$$\mathrm{var}(\hat{G}_{ii}) \geq \frac{1}{L(\bar{\eta}_i - 1)}, \tag{97}$$

$$\mathbf{R}_{\boldsymbol{\vartheta}} = (\mathbf{M}_1 - \mathbf{M}_2\mathbf{M}_1^{-*}\mathbf{M}_2^*)^* = (\mathbf{M}_1 - \mathbf{M}_2^*\mathbf{M}_1^{-1}\mathbf{M}_2)$$
$$= \sum_{i=1}^N (\eta_i - 1)\mathbf{L}_{ii}\otimes\mathbf{L}_{ii} + \sum_{i=1}^N\sum_{\substack{j=1\\j\neq i}}^N \kappa_i\mathbf{L}_{ii}\otimes\mathbf{L}_{jj} - \left(\sum_{i=1}^N(\beta_i-1)\mathbf{L}_{ii}\otimes\mathbf{L}_{ii} + \sum_{i=1}^N\sum_{\substack{j=1\\j\neq i}}^N \gamma_j\xi_i\mathbf{L}_{ii}\otimes\mathbf{L}_{jj} + \sum_{i=1}^N\sum_{\substack{j=1\\j\neq i}}^N \mathbf{L}_{ij}\otimes\mathbf{L}_{ji}\right)^*$$
$$\cdot\left(\sum_{i=1}^N(\eta_i-1)^{-1}\mathbf{L}_{ii}\otimes\mathbf{L}_{ii} + \sum_{i=1}^N\sum_{\substack{j=1\\j\neq i}}^N \kappa_i^{-1}\mathbf{L}_{ii}\otimes\mathbf{L}_{jj}\right)\left(\sum_{i=1}^N(\beta_i-1)\mathbf{L}_{ii}\otimes\mathbf{L}_{ii} + \sum_{i=1}^N\sum_{\substack{j=1\\j\neq i}}^N \gamma_j\xi_i\mathbf{L}_{ii}\otimes\mathbf{L}_{jj} + \sum_{i=1}^N\sum_{\substack{j=1\\j\neq i}}^N \mathbf{L}_{ij}\otimes\mathbf{L}_{ji}\right) \tag{88}$$

$$\text{var}(\hat{G}_{ij}) \geq \frac{1}{L}\frac{\bar{\kappa}_j}{\bar{\kappa}_i\bar{\kappa}_j - 1}. \tag{98}$$

For temporally non-iid sources, we collect L samples of source i in a row vector $\bar{\mathbf{s}}_i$ with pdf $p(\bar{\mathbf{s}}_i)$. The iCRB for G_{ij} then changes to

$$\text{var}(\hat{G}_{ij}) \geq \frac{1}{L}\frac{\bar{\rho}_{ji}}{\bar{\rho}_{ij}\bar{\rho}_{ji} - 1}. \tag{99}$$

$\bar{\eta}_i$ and $\bar{\rho}_{ij}$ are defined as [3]

$$\bar{\eta}_i - 1 = L\left(\text{E}\left[\frac{1}{L}\bar{\mathbf{s}}_i\bar{\boldsymbol{\varphi}}_i^T(\bar{\mathbf{s}}_i)\right]^2 - 1\right), \tag{100}$$

$$\bar{\rho}_{ij} = \frac{1}{L}\,\text{tr}(\bar{\mathbf{R}}_j\bar{\boldsymbol{\kappa}}_i), \tag{101}$$

where $\bar{\mathbf{R}}_j = \text{E}[\bar{\mathbf{s}}_j^T\bar{\mathbf{s}}_j]$ and $\bar{\boldsymbol{\kappa}}_i = \text{E}[\bar{\boldsymbol{\varphi}}_i^T(\bar{\mathbf{s}}_i)\bar{\boldsymbol{\varphi}}_i(\bar{\mathbf{s}}_i)]$. $\bar{\boldsymbol{\varphi}}_i(\mathbf{s}_i)$ is a row vector whose l-th element is defined as

$$[\bar{\boldsymbol{\varphi}}_i(\bar{\mathbf{s}}_i)]_l = -\frac{\partial}{\partial\bar{s}_i(l)}\ln p(\bar{\mathbf{s}}_i). \tag{102}$$

For Gaussian sources, $\bar{\eta}_i = 3$ and $\bar{\rho}_{ij} = \frac{1}{L}\,\text{tr}(\bar{\mathbf{R}}_j\bar{\mathbf{R}}_i^{-1})$ [3], [34].

Acknowledgement

The authors would like to thank the anonymous reviewers for their helpful and valuable comments and suggestions.

References

[1] P. Comon, "Independent component analysis, a new concept?" *Signal Process.*, vol. 36, no. 3, pp. 287–314, April 1994.
[2] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*. John Wiley & Sons, 2001.
[3] P. Comon and C. Jutten, Eds., *Handbook of blind source separation : independent component analysis and applications*, 1st ed. Amsterdam: Elsevier, 2010.
[4] W. Wirtinger, "Zur formalen Theorie der Funktionen von mehr komplexen Veränderlichen," *Mathematische Annalen*, vol. 97, no. 1, pp. 357–375, 1927.
[5] J. Eriksson and V. Koivunen, "Complex random vectors and ICA models: identifiability, uniqueness, and separability," *IEEE Transactions on Information Theory*, vol. 52, no. 3, pp. 1017 –1029, March 2006.
[6] S. Fiori, "On blind separation of complex-valued sources by extended hebbian learning," *IEEE Signal Processing Letters*, vol. 8, no. 8, pp. 217 –220, Aug. 2001.
[7] ——, "Neural independent component analysis by maximum-mismatch learning principle," *Neural Networks*, vol. 16, no. 8, pp. 1201 – 1221, 2003.
[8] ——, "Nonlinear complex-valued extensions of hebbian learning: An essay," *Neural Comput.*, vol. 17, no. 4, pp. 779–838, Apr. 2005.
[9] T. Adali, P. Schreier, and L. Scharf, "Complex-valued signal processing: The proper way to deal with impropriety," *IEEE Transactions on Signal Processing*, vol. 59, no. 11, pp. 5101 –5125, nov. 2011.
[10] P. Tichavsky, Z. Koldovsky, and E. Oja, "Performance analysis of the FastICA algorithm and Cramér-Rao bounds for linear independent component analysis," *IEEE Transactions on Signal Processing*, vol. 54, no. 4, Apr. 2006.
[11] E. Ollila, H.-J. Kim, and V. Koivunen, "Compact Cramér-Rao bound expression for independent component analysis," *IEEE Transactions on Signal Processing*, vol. 56, no. 4, Apr. 2008.
[12] J. Cardoso and A. Souloumiac, "Blind beamforming for non-gaussian signals," *Radar and Signal Processing, IEE Proceedings F*, vol. 140, no. 6, pp. 362 –370, Dec. 1993.
[13] L. De Lathauwer and B. De Moor, "On the blind separation of non-circular sources," in *Proc. European Signal Processing Conference (EUSIPCO)*, Toulouse, France, Sept. 2002.
[14] J. Eriksson and V. Koivunen, "Complex-valued ICA using second order statistics," in *Proc. IEEE Workshop on Machine Learning (MLSP)*, 2004, pp. 183 –192.
[15] J.-F. Cardoso and T. Adali, "The maximum likelihood approach to complex ICA," in *Proc. IEEE Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 5, May 2006.
[16] M. Novey and T. Adali, "Adaptable nonlinearity for complex maximization of nongaussianity and a fixed-point algorithm," *Proc. IEEE Workshop on Machine Learning for Signal Processing (MLSP)*, Sept. 2006.
[17] S. C. Douglas, "Fixed-point algorithms for the blind separation of arbitrary complex-valued non-gaussian signal mixtures," *EURASIP J. Appl. Signal Process.*, vol. 2007, no. 1, January 2007.
[18] M. Novey and T. Adali, "On extending the complex FastICA algorithm to noncircular sources," *IEEE Transactions on Signal Processing*, vol. 56, no. 5, pp. 2148 –2154, May 2008.
[19] H. Li and T. Adali, "Algorithms for complex ML ICA and their stability analysis using Wirtinger calculus," *IEEE Transactions on Signal Processing*, vol. 58, no. 12, Dec. 2010.
[20] X.-L. Li and T. Adali, "Complex independent component analysis by entropy bound minimization," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 57, no. 7, pp. 1417 –1430, July 2010.
[21] A. Yeredor, "Performance analysis of the strong uncorrelating transformation in blind separation of complex-valued sources," *IEEE Transactions on Signal Processing*, vol. 60, no. 1, pp. 478–483, Jan. 2012.
[22] T. Adali, H. Li, M. Novey, and J.-F. Cardoso, "Complex ICA using nonlinear functions," *IEEE Transactions on Signal Processing*, vol. 56, no. 9, pp. 4536 –4544, Sept. 2008.
[23] B. Loesch and B. Yang, "Cramér-Rao bound for circular complex independent component analysis," *Proc. International Conference on Latent Variable Analysis and Signal Separation (LVA/ICA)*, 2012.
[24] A. Hjørungnes, *Complex-Valued Matrix Derivatives*. Cambridge University Press, 2011.
[25] P. J. Schreier and L. L. Scharf, *Statistical signal processing of complex-valued data: The theory of improper and noncircular signals*. Cambridge: Univ. Press, 2010.
[26] D. H. Brandwood, "A complex gradient operator and its application in adaptive array theory," *IEE Proc.*, vol. 130, pp. 11–16, 1983.
[27] R. Remmert, *Theory of complex functions*, ser. Graduate texts in mathematics. Springer-Verlag, 1991.
[28] E. Ollila, V. Koivunen, and J. Eriksson, "On the Cramér-Rao bound for the constrained and unconstrained complex parameters," *IEEE Workshop on Sensor Array and Multichannel Signal Processing*, 2008.

[29] R. A. Horn and C. R. Johnson, *Matrix analysis*, 1st ed. Cambridge: University Press, 1999.
[30] A. Jagannatham and B. Rao, "Cramér-Rao lower bound for constrained complex parameters," *IEEE Signal Processing Letters*, vol. 11, no. 11, pp. 875 – 878, Nov. 2004.
[31] T. Adali and H. Li, "Complex-valued adaptive signal processing, ch. 1," in *Adaptive Signal Processing: Next Generation Solutions*, T. Adali and S. Haykin, Eds. John Wiley & Sons, Inc., 2010, pp. 1–85.
[32] E. Doron, A. Yeredor, and P. Tichavsky, "Cramér-Rao-induced bound for blind separation of stationary parametric gaussian sources," *ieespl*, vol. 14, no. 6, pp. 417 –420, june 2007.
[33] M. Anderson, X.-L. Li, P. A. Rodriquez, and T. Adali, "An effective decoupling method for matrix optimization and its application to the ICA problem," *Proc. IEEE Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2012.
[34] A. Yeredor, "Blind separation of gaussian sources with general covariance structures: Bounds and optimal estimation," *IEEE Transactions on Signal Processing*, vol. 58, no. 10, pp. 5057 –5068, oct. 2010.
[35] M. Novey, T. Adali, and A. Roy, "A complex generalized gaussian distribution – characterization, generation, and estimation," *IEEE Transactions on Signal Processing*, vol. 58, no. 3, pp. 1427 –1433, March 2010.
[36] S.-I. Amari, "Natural gradient works efficiently in learning," *Neural Computation*, vol. 10, no. 2, pp. 251–276, 1998.
[37] S. Fiori, "Lie-group-type neural system learning by manifold retractions," *Neural Networks*, vol. 21, no. 10, pp. 1524 – 1529, 2008.
[38] S. Fiori and T. Tanaka, "An algorithm to compute averages on matrix Lie groups," *IEEE Transactions on Signal Processing*, vol. 57, no. 12, pp. 4734–4743, Dec. 2009.
[39] B. Picinbono, "Second-order complex random vectors and normal distributions," *IEEE Transactions on Signal Processing*, vol. 44, no. 10, pp. 2637 –2640, Oct. 1996.
[40] X.-L. Li and T. Adali, "Blind separation of noncircular correlated sources using gaussian entropy rate," *IEEE Transactions on Signal Processing*, vol. 59, no. 6, pp. 2969 –2975, june 2011.

loesch

Benedikt Loesch received the M.Sc. degree from Georgia Institute of Technology, USA in 2007 and the Dipl.-Ing. degree from the University of Stuttgart, Germany in 2008, both in electrical engineering. Simultaneously, he studied music at the University of Music and Performing Arts, Stuttgart and received diploma degrees in music theory/new media, violin education and violin performance. From 2008 to 2012 he was a research assistant at the Institute of Signal Processing and System Theory, raphy University of Stuttgart. In this time he worked in the area of statistical signal processing, focusing especially on blind source separation. He recently joined Robert Bosch GmbH, Germany where he works on signal processing and classification for automotive safety systems.

yang

Bin Yang (SM '06) received the Dipl.-Ing. and Ph.D. degree in 1986 and 1991 from the Ruhr University Bochum, Germany, all in electrical engineering. From 1996 to 2001, he was a senior researcher on mobile communications at Infineon Technologies, Germany. Since 2001 he is Professor and head of the Institute of Signal Processing and System Theory at University of Stuttgart, Germany. His research interests include theory and algorithms of statistical signal processing and various applications like array processing, automotive safety systems, localization, blind methods, emotion recognition etc.