Chapter 3
Performance Study for Complex Independent Component Analysis

Benedikt Loesch and Bin Yang

Abstract The goal of independent component analysis (ICA) is to decompose observed signals into components as independent as possible. In linear instantaneous blind source separation, ICA is used to separate linear instantaneous mixtures of source signals into signals that are as close as possible to the original signals. In the estimation of the so-called demixing matrix one has to distinguish two different factors:

1. Variance of the estimated inverse mixing matrix in the noiseless case due to randomness of the sources.
2. Bias of the demixing matrix from the inverse mixing matrix:

This chapter studies both factors for circular and noncircular complex mixtures. It is important to note that the complex case is not directly equivalent to the real case of twice larger dimension. In the derivations, we aim to clearly show the connections and differences between the complex and real cases. In the first part of the chapter, we derive a closed-form expression for the CRB of the demixing matrix for instantaneous noncircular complex mixtures. We also study the CRB numerically for the family of noncircular complex generalized Gaussian distributions (GGD) and compare it to simulation results of several ICA estimators. In the second part, we consider a linear noisy noncircular complex mixing model and derive an analytic expression for the

B. Loesch (✉) · B. Yang
Institute of Signal Processing and System Theory, University of Stuttgart, Stuttgart, Germany
e-mail: benedikt.loesch@iss.uni-stuttgart.de

B. Yang
e-mail: bin.yang@iss.uni-stuttgart.de

demixing matrix of ICA based on the Kullback-Leibler divergence (KLD). We show that for a wide range of both the shape parameter and the noncircularity index of the GGD, the signal-to-interference-plus-noise ratio (SINR) of KLD-based ICA is close to that of linear MMSE estimation. Furthermore, we show how to extend our derivations to the overdetermined case ($M > N$) with circular complex noise.

3.1 Introduction

The goal of independent component analysis (ICA) is to decompose observed signals into components as independent as possible. In linear instantaneous blind source separation, ICA is used to separate linear instaneous mixtures of source signals into signals which are as close as possible to the original signals. In the estimation of the so-called demixing matrix, one has to distinguish two different factors:

1. Variance of the estimated inverse mixing matrix in the noiseless case due to randomness of the sources. This variance can be lower bounded by the Cramér-Rao bound for ICA derived for the real case in [41, 45] and for the circular and noncircular complex case in [33, 35].
2. Bias of the demixing matrix from the inverse mixing matrix: As already noted in [16], the presence of noise leads to a bias of the demixing matrix from the inverse mixing matrix. Often a bias of an estimator is considered to be unwanted, but in the case of noisy ICA the bias of the demixing matrix from the inverse mixing matrix actually leads to a reduced noise level in the separated signals and hence it can be considered to be desired.

This chapter studies both factors for circular and noncircular complex mixtures. It is important to note that the complex case is not directly equivalent to the real case of twice larger dimension [19]. In the derivations, we aim to clearly show the connections and differences between the complex and real cases.

In many practical applications such as audio processing in frequency-domain or telecommunication, the signals are complex. While many publications focus on circular[1] complex signals (as traditionally assumed in signal processing), [4, 36, 44] provide a good overview of applications with noncircular complex signals and discuss how to properly deal with noncircularity. Many signals of practical interest are noncircular. Digital modulation schemes[2] usually produce noncircular complex baseband signals, since the symbol constellations in the complex plane are only rotationally symmetric for a discrete set of rotation angles but not any arbitrary real rotation angle as necessary for circularity [2]. Another source of noncircularity is an imbalance between the in-phase and quadrature (I/Q) components of communication signals. Noncircularity can also be exploited in feature extraction in electrocardio-

---

[1] See Sect. 3.1.1 for a definition.

[2] Examples of digital modulation schemes are phase shift keying (PSK), pulse amplitude modulation (PAM) or quadrature amplitude modulation (QAM).

grams (ECGs) and in the analysis of functional magnetic resonance imaging (fMRI) [4]. Moreover, the theory of noncircularity has found applications in acoustics and optics [44].

Although a large number of different algorithms for complex ICA have been proposed [7, 10, 14, 17, 18, 20, 29, 30, 38, 39], the CRB for the complex demixing matrix has only been derived recently in [33, 35]. General conditions regarding identifiability, uniqueness, and separability for complex ICA can be found in Eriksson and Koivunen [19]. Yeredor [48] provides a performance analysis 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 [48] do not apply for ICA algorithms exploiting also the non-Gaussianity of the sources. As discussed in [3, 4], many ICA approaches exploiting non-Gaussianity of the sources are intimately related and can be studied under the umbrella of a maximum likelihood framework whose asymptotic performance reaches the CRB if the assumed distribution of the sources matches the true distribution.

The structure of the separation problem changes substantially if we account for additive noise. As discussed in [12], the mixing model is no longer equivariant and the likelihood contrast can no longer be assimilated to mutual information. Furthermore, the ML estimate of the source signals is no longer a linear function of the observations [23]. Source estimation from noisy mixtures can be classified into linear and nonlinear separation. In linear ICA, the presence of noise leads to a bias in the estimation of the mixing matrix. Douglas et al. [16] introduced measures to reduce this bias. Cardoso [8] showed that the performance of noisy source separation depends on the distribution of the sources, the signal-to-noise ratio (SNR) and the mixing matrix. Davies [13] showed for the real case that it is not meaningful to estimate both the mixing matrix and the full covariance matrix of the noise from the data. Koldovsky and Tichavsky [27, 28] drew parallels between linear minimum mean squared error (MMSE) estimation and ICA for the real data case. Up to now, closed-form expressions for the bias of the ICA solution in the complex case have not been derived except for the recent work of Loesch and Yang [34].

After a review of notation for complex-valued signals, complex ICA, and the CRB for a complex parameter vector in Sect. 3.1.1, we derive a closed-form expression for the CRB of the demixing matrix for instantaneous noncircular complex mixtures in Sect. 3.2. We first introduce the signal model and the assumptions in Sect. 3.2.1 and then derive the CRB for the complex demixing matrix in Sect. 3.2.2. Section 3.2.3 discusses the circular complex case and noncircular complex Gaussian case as two special cases of the CRB. In Sect. 3.2.4, we study the CRB numerically for family of noncircular complex generalized Gaussian distributions[3] (GGD) and compare it to simulation results of several ICA estimators.

In Sect. 3.3, we consider a linear noisy noncircular complex mixing model and derive an analytic expression for the demixing matrix of ICA based on the Kullback-Leibler divergence (KLD) [34]. This expression contains the circular complex and

---

[3] See Sect. 3.2.4 for a definition.

real case as special cases. The derivation is done using a perturbation analysis valid for small noise variance.[4] In Sect. 3.3.3, we show that for a wide range of both the shape parameter and the noncircularity index of the GGD, the signal-to-interference-plus-noise ratio (SINR) of KLD-based ICA is close to that of linear MMSE estimation. We also discuss the situations where the two solutions differ. Furthermore, we extend our derivations to the overdetermined case ($M > N$) with circular complex noise in Sect. 3.3.4.

Compared to our previous journal and conference publications [33–35], we extend the performance study to a larger number of ICA algorithms and extend the results for noisy mixtures to the overdetermined case.

3.1.1 Notations for Complex-Valued Signals

3.1.1.1 Complex Random Vector

Let $\mathbf{x} = \mathbf{x}_R + j\mathbf{x}_I \in \mathbb{C}^N$ be a complex random *column* 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

$$\mathrm{cov}(\mathbf{x}) = \mathrm{E}\left[(\mathbf{x} - \mathrm{E}[\mathbf{x}])(\mathbf{x} - \mathrm{E}[\mathbf{x}])^H\right]. \tag{3.1}$$

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

$$\mathrm{pcov}(\mathbf{x}) = \mathrm{E}\left[(\mathbf{x} - \mathrm{E}[\mathbf{x}])(\mathbf{x} - \mathrm{E}[\mathbf{x}])^T\right]. \tag{3.2}$$

$(\cdot)^T$ and $(\cdot)^H$ stand for transpose and complex conjugate transpose of a vector or matrix. The augmented covariance matrix of $\mathbf{x}$ is the covariance matrix of the augmented vector $\underline{\mathbf{x}} = \left[\mathbf{x}^T\ \mathbf{x}^H\right]^T$:

$$\mathrm{cov}(\underline{\mathbf{x}}) = \begin{bmatrix} \mathrm{cov}(\mathbf{x}) & \mathrm{pcov}(\mathbf{x}) \\ \mathrm{pcov}(\mathbf{x})^* & \mathrm{cov}(\mathbf{x})^* \end{bmatrix}. \tag{3.3}$$

$\mathbf{x}$ 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 $\gamma = \mathrm{E}\left[s^2\right] = 0$. There exist noncircular complex random variables with $\gamma = 0$. For simplicity, however,

---

[4] For a large noise variance σ^2 the theoretical analysis cannot fully describe the behavior of KLD-based ICA since we only take into account terms of order σ^2. However, simulation results show that KLD-based ICA still performs similarly to linear MMSE estimation.

we use the second-order noncircularity index $\gamma = \mathrm{E}\left[s^2\right]$ to quantify noncircularity in the remainder of this chapter.

3.1.1.2 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 [21, 44]. The complex gradient vectors $\nabla_{\boldsymbol{\theta}} f$ and $\nabla_{\boldsymbol{\theta}^*} f$ are given by

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

$$\nabla_{\boldsymbol{\theta}^*} f = \frac{\partial f}{\partial \boldsymbol{\theta}^*} = \frac{1}{2}\left(\frac{\partial \tilde{f}}{\partial \boldsymbol{\theta}_R} + j\frac{\partial \tilde{f}}{\partial \boldsymbol{\theta}_I}\right) \in \mathbb{C}^M. \tag{3.4}$$

The stationary points of $f(\cdot)$ and $\tilde{f}(\cdot)$ are given by $\left(\frac{\partial \tilde{f}}{\partial \boldsymbol{\theta}_R} = \mathbf{0} \text{ and } \frac{\partial \tilde{f}}{\partial \boldsymbol{\theta}_I} = \mathbf{0}\right)$ 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 $-\frac{\partial f}{\partial \boldsymbol{\theta}}$ [6]. 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 [43]. Furthermore, we note that the Wirtinger derivatives defined in (3.4) are also valid for partial derivatives of f with respect to a parameter *matrix* $\boldsymbol{\Theta}$. In this chapter, we will also use real derivatives which we denote as $(\cdot)'$ wherever possible.

3.1.1.3 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 $\bar{\boldsymbol{\theta}} = \left[\boldsymbol{\theta}_R^T \ \boldsymbol{\theta}_I^T\right]^T$:

$$\mathrm{cov}(\bar{\boldsymbol{\theta}}) = \begin{bmatrix} \mathrm{cov}(\boldsymbol{\theta}_R) & \mathrm{cov}(\boldsymbol{\theta}_R, \boldsymbol{\theta}_I) \\ \mathrm{cov}(\boldsymbol{\theta}_I, \boldsymbol{\theta}_R) & \mathrm{cov}(\boldsymbol{\theta}_I) \end{bmatrix} \geq L^{-1}\mathbf{J}_{\bar{\boldsymbol{\theta}}}^{-1}, \tag{3.5}$$

where $\mathrm{cov}(\mathbf{x}, \mathbf{y}) = \mathrm{E}\left[(\mathbf{x} - \mathrm{E}[\mathbf{x}])(\mathbf{y} - \mathrm{E}[\mathbf{y}])^T\right]$ denotes the cross-covariance matrix 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

$$\mathscr{J}_{\boldsymbol{\theta}} = \begin{bmatrix} \mathscr{I}_{\boldsymbol{\theta}} & \mathscr{P}_{\boldsymbol{\theta}} \\ \mathscr{P}_{\boldsymbol{\theta}}^* & \mathscr{I}_{\boldsymbol{\theta}}^* \end{bmatrix}, \tag{3.6}$$

where $\mathscr{I}_{\boldsymbol{\theta}} = \mathrm{E}\left[\left\{\nabla_{\boldsymbol{\theta}^*} \ln p(\mathbf{x}; \boldsymbol{\theta})\right\}\left\{\nabla_{\boldsymbol{\theta}^*} \ln p(\mathbf{x}; \boldsymbol{\theta})\right\}^H\right]$ is called the information matrix and $\mathscr{P}_{\boldsymbol{\theta}} = \mathrm{E}\left[\left\{\nabla_{\boldsymbol{\theta}^*} \ln p(\mathbf{x}; \boldsymbol{\theta})\right\}\left\{\nabla_{\boldsymbol{\theta}^*} \ln p(\mathbf{x}; \boldsymbol{\theta})\right\}^T\right]$ the 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 $\boldsymbol{\theta}$ [42, 44]

$$\begin{bmatrix} \mathrm{cov}(\hat{\boldsymbol{\theta}}) & \mathrm{pcov}(\hat{\boldsymbol{\theta}}) \\ \mathrm{pcov}(\hat{\boldsymbol{\theta}})^* & \mathrm{cov}(\hat{\boldsymbol{\theta}})^* \end{bmatrix} \geq \left(L \mathscr{J}_{\boldsymbol{\theta}}\right)^{-1} = L^{-1} \begin{bmatrix} \mathscr{I}_{\boldsymbol{\theta}} & \mathscr{P}_{\boldsymbol{\theta}} \\ \mathscr{P}_{\boldsymbol{\theta}}^* & \mathscr{I}_{\boldsymbol{\theta}}^* \end{bmatrix}^{-1}. \tag{3.7}$$

Note that the complex CRB (3.7) can be transformed to the corresponding real CRB (3.5) by using the transform $\mathbf{J}_{\bar{\boldsymbol{\theta}}}^{-1} = \frac{1}{2} \mathbf{T} \mathscr{J}_{\boldsymbol{\theta}}^{-1} \mathbf{T}^{-1}$ [42], 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 [22], we get from (3.7)

$$\begin{bmatrix} \mathrm{cov}(\hat{\boldsymbol{\theta}}) & \mathrm{pcov}(\hat{\boldsymbol{\theta}}) \\ \mathrm{pcov}(\hat{\boldsymbol{\theta}})^* & \mathrm{cov}(\hat{\boldsymbol{\theta}})^* \end{bmatrix} \geq L^{-1} \begin{bmatrix} \mathbf{R}_{\boldsymbol{\theta}}^{-1} & -\mathbf{R}_{\boldsymbol{\theta}}^{-1} \mathbf{Q}_{\boldsymbol{\theta}} \\ -\mathbf{Q}_{\boldsymbol{\theta}}^H \mathbf{R}_{\boldsymbol{\theta}}^{-1} & \mathbf{R}_{\boldsymbol{\theta}}^{-*} \end{bmatrix} \tag{3.8}$$

with $\mathbf{R}_{\boldsymbol{\theta}} = \mathscr{I}_{\boldsymbol{\theta}} - \mathscr{P}_{\boldsymbol{\theta}} \mathscr{I}_{\boldsymbol{\theta}}^{-*} \mathscr{P}_{\boldsymbol{\theta}}^*$ and $\mathbf{Q}_{\boldsymbol{\theta}} = \mathscr{P}_{\boldsymbol{\theta}} \mathscr{I}_{\boldsymbol{\theta}}^{-*}$. $\mathbf{A}^{-*}$ is a short notation for $\left(\mathbf{A}^{-1}\right)^* = (\mathbf{A}^*)^{-1}$. Often we are interested in the bound for $\mathrm{cov}(\hat{\boldsymbol{\theta}})$ only, which can be obtained from (3.8) as

$$\mathrm{cov}(\hat{\boldsymbol{\theta}}) \geq L^{-1} \mathbf{R}_{\boldsymbol{\theta}}^{-1} = L^{-1} (\mathscr{I}_{\boldsymbol{\theta}} - \mathscr{P}_{\boldsymbol{\theta}} \mathscr{I}_{\boldsymbol{\theta}}^{-*} \mathscr{P}_{\boldsymbol{\theta}}^*)^{-1}. \tag{3.9}$$

Note that (3.9) gives a bound solely on the covariance matrix of an unbiased estimator. If an estimator reaches that bound, i.e., $\mathrm{cov}(\hat{\boldsymbol{\theta}}) = L^{-1} \mathbf{R}_{\boldsymbol{\theta}}^{-1}$, it does not imply that it also reaches the general CRB defined in (3.7). Only if the pseudo-information matrix $\mathscr{P}_{\boldsymbol{\theta}}$ vanishes, $\mathrm{cov}(\hat{\boldsymbol{\theta}}) = L^{-1} \mathbf{R}_{\boldsymbol{\theta}}^{-1}$ implies that $\hat{\boldsymbol{\theta}}$ reaches the CRB (3.7).

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 either [42] or [24]. If the unconstrained Fisher information matrix is singular, only the constrained CRB from [24] can be applied.

3.2 Cramér-Rao Bound for Complex ICA

For the performance analysis of ICA algorithms, it is useful to have a lower bound for the covariance matrix of estimators of 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 [41, 45] which we summarized in Appendix 1. However, in many practical applications such as audio processing in frequency-domain or telecommunication, the signals are complex and hence we need the CRB for complex ICA.

3.2.1 Signal Model and Assumptions

Throughout this section, we assume an instantaneous complex linear square noiseless mixing model

$$\mathbf{x} = \mathbf{As} \tag{3.10}$$

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. $\mathbf{s} = [s_1, \ldots, s_N]^T \in \mathbb{C}^N$ are N independent random variables with zero mean, unit variance $\mathrm{E}\left[|s_i|^2\right] = 1$ and second-order noncircularity index $\gamma_i = E[s_i^2] \in [0, 1]$.[5] 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) $p_{s_i}(s_i)$ of different source signals s_i can be identical or different. $p_{s_i}(s_i)$ is continuously differentiable with respect to s_i and s_i^* in the sense of Wirtinger derivatives [46] which have been shortly reviewed in Sect. 3.1.1. All required expectations exist.

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

$$\mathbf{y} = \mathbf{Wx} = \mathbf{WAs} \tag{3.11}$$

such that $\mathbf{y}$ is "as close to $\mathbf{s}$" as possible according to some metric.

The ideal solution for $\mathbf{W}$ is $\mathbf{A}^{-1}$, neglecting scaling, phase, and permutation ambiguity [19]. If we know the pdfs $p_{s_i}(s_i)$ perfectly, there is no scaling ambiguity. Due to the "working" assumption $\gamma_i \in [0, 1]$ (see Appendix 2), there is no phase ambiguity for noncircular sources ($\gamma_i > 0$) [1, 37]. A phase ambiguity occurs only for circular sources ($\gamma_i = 0$). Noncircular sources s_i which do not comply with the assumption $\gamma_i \in [0, 1]$ can be transformed according to $s_i e^{j\alpha_i}$ such that $\gamma_i \in [0, 1]$.

---

[5] Due to the inherent scaling ambiguity between the mixing matrix $\mathbf{A}$ and the source signals $\mathbf{s}$, without loss of generality, we can scale $\mathbf{s}$ and accordingly $\mathbf{A}$ such that $\mathrm{E}\left[|s_i|^2\right] = 1$ and $\gamma_i \in [0, 1]$.

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 different time instants $t_1 \neq t_2$,
- nonstationarity, i.e., the statistical properties of $s(t)$ change over time.

In this section, we focus on noncircular complex source signals with independent and identically distributed (iid) time samples. An extension to temporally non-iid sources, i.e., to incorporate nonstationarity and nonwhiteness of the sources, has been given in [35].

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 [19].

3.2.2 Derivation of the Cramér-Rao Bound

We form the parameter vector

$$\boldsymbol{\theta} = \mathrm{vec}(\mathbf{W}^T) = [\mathbf{w}_1^T, \ldots, \mathbf{w}_N^T]^T \in \mathbb{C}^{N^2} \tag{3.12}$$

where $\mathbf{w}_i^T$ denotes the i-th row vector of $\mathbf{W}$. The $\mathrm{vec}(\cdot)$ operator stacks the columns of its argument into one long column vector. Given the pdfs $p_{s_i}(s_i)$ of the complex source signals s_i and the complex linear transform $\mathbf{x} = \mathbf{As}$, it is easy to derive the pdf of $\mathbf{x}$ as $p(\mathbf{x}; \boldsymbol{\theta}) = |\det(\mathbf{W})|^2 \prod_{i=1}^{N} p_{s_i}(\mathbf{w}_i^T \mathbf{x})$. 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 [2, 3, 21], we obtain

$$\frac{\partial}{\partial \mathbf{W}^H} \ln p(\mathbf{x}; \boldsymbol{\theta}) = \mathbf{A}^* - \mathbf{x}^* \boldsymbol{\varphi}^T(\mathbf{Wx}) = \mathbf{A}^*(\mathbf{I} - \mathbf{s}\boldsymbol{\varphi}^H(\mathbf{s}))^* \tag{3.13}$$

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

$$\varphi_i(s_i) = -\frac{\partial}{\partial s_i^*} \ln p_{s_i}(s_i) = -\frac{1}{2} \frac{1}{p_{s_i}(s_i)} \left[\frac{\partial p_{s_i}(s_i)}{\partial s_{i,R}} + j \frac{\partial p_{s_i}(s_i)}{\partial s_{i,I}} \right]. \tag{3.14}$$

Since $\boldsymbol{\theta} = \mathrm{vec}(\mathbf{W}^T)$, we get

$$\nabla_{\boldsymbol{\theta}^*} \ln p_{\mathbf{x}}(\mathbf{x}; \boldsymbol{\theta}) = \mathrm{vec}\left(\frac{\partial}{\partial \mathbf{W}^H} \ln p_{\mathbf{x}}(\mathbf{x}; \boldsymbol{\theta}) \right) = \left[(\mathbf{I} \otimes \mathbf{A}) \, \mathrm{vec}\left(\mathbf{I} - \mathbf{s}\boldsymbol{\varphi}(\mathbf{s})^H \right) \right]^*, \tag{3.15}$$

where $\mathbf{A} \otimes \mathbf{B} = \begin{bmatrix} a_{ij}\mathbf{B} \end{bmatrix}$ denotes the Kronecker product of $\mathbf{A}$ and $\mathbf{B}$. Hence, the information and pseudo-information matrix in (3.6) become

$$
\begin{aligned}
\mathscr{I}_{\boldsymbol{\theta}} &= \left((\mathbf{I} \otimes \mathbf{A})\mathrm{E}\left[\mathrm{vec}\{\mathbf{I} - \mathbf{s}\boldsymbol{\varphi}^H(\mathbf{s})\} \, \mathrm{vec}\{\mathbf{I} - \mathbf{s}\boldsymbol{\varphi}^H(\mathbf{s})\}^H \right](\mathbf{I} \otimes \mathbf{A}^H) \right)^* \\
&= \left((\mathbf{I} \otimes \mathbf{A})\mathbf{M}_1(\mathbf{I} \otimes \mathbf{A}^H) \right)^*, \quad (3.16) \\
\mathscr{P}_{\boldsymbol{\theta}} &= \left((\mathbf{I} \otimes \mathbf{A})\mathrm{E}\left[\mathrm{vec}\{\mathbf{I} - \mathbf{s}\boldsymbol{\varphi}^H(\mathbf{s})\} \, \mathrm{vec}\{\mathbf{I} - \mathbf{s}\boldsymbol{\varphi}^H(\mathbf{s})\}^T \right](\mathbf{I} \otimes \mathbf{A}^T) \right)^* \\
&= \left((\mathbf{I} \otimes \mathbf{A})\mathbf{M}_2(\mathbf{I} \otimes \mathbf{A}^T) \right)^*, \quad (3.17)
\end{aligned}
$$

where

$$
\begin{aligned}
\mathbf{M}_1 &= \mathrm{E}\left[\mathrm{vec}\{\mathbf{I} - \mathbf{s}\boldsymbol{\varphi}^H(\mathbf{s})\} \, \mathrm{vec}\{\mathbf{I} - \mathbf{s}\boldsymbol{\varphi}^H(\mathbf{s})\}^H \right] \\
\text{and} \quad \mathbf{M}_2 &= \mathrm{E}\left[\mathrm{vec}\{\mathbf{I} - \mathbf{s}\boldsymbol{\varphi}^H(\mathbf{s})\} \, \mathrm{vec}\{\mathbf{I} - \mathbf{s}\boldsymbol{\varphi}^H(\mathbf{s})\}^T \right]. \quad (3.18)
\end{aligned}
$$

3.2.2.1 Induced CRB for the Gain Matrix G = WA

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 $\mathbf{G}$.[6] 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

$$
\boldsymbol{\vartheta} = \mathrm{vec}(\mathbf{G}^T) = \mathrm{vec}(\mathbf{A}^T\mathbf{W}^T) = (\mathbf{I} \otimes \mathbf{A}^T)\,\mathrm{vec}(\mathbf{W}^T) = (\mathbf{I} \otimes \mathbf{A}^T)\boldsymbol{\theta}. \quad (3.19)
$$

The covariance matrix of $\hat{\boldsymbol{\vartheta}} = \mathrm{vec}((\hat{\mathbf{W}}\mathbf{A})^T)$ is given by $\mathrm{cov}(\hat{\boldsymbol{\vartheta}}) = (\mathbf{I} \otimes \mathbf{A}^T)\,\mathrm{cov}(\hat{\boldsymbol{\theta}})$ $(\mathbf{I} \otimes \mathbf{A}^*)$ where $\hat{\boldsymbol{\theta}} = \mathrm{vec}(\hat{\mathbf{W}}^T)$. By combining (3.9) with (3.16) and (3.17), we get

$$
\mathrm{cov}(\hat{\boldsymbol{\vartheta}}) \geq L^{-1}(\mathbf{I} \otimes \mathbf{A}^T)(\mathscr{I}_{\boldsymbol{\theta}} - \mathscr{P}_{\boldsymbol{\theta}}\mathscr{I}_{\boldsymbol{\theta}}^{-*}\mathscr{P}_{\boldsymbol{\theta}}^*)^{-1}(\mathbf{I} \otimes \mathbf{A}^*) = L^{-1}\mathbf{R}_{\boldsymbol{\vartheta}}^{-1} \quad (3.20)
$$

---

[6] Some authors [5, 15, 47] prefer the so-called expected interference-to-source ratio (ISR) matrix whose elements $\overline{\mathrm{ISR}}_{ij}$ are defined (for $i \neq j$ and unit variance sources) as $\overline{\mathrm{ISR}}_{ij} = \mathrm{E}\left[\frac{|G_{ij}|^2}{|G_{ii}|^2} \right]$, where G_{ii} denotes the diagonal elements and G_{ij} the off-diagonal elements of $\mathbf{G}$. To compute $\overline{\mathrm{ISR}}_{ij}$, usually $G_{ii} \approx 1$ (i.e., $\mathrm{var}(G_{ii}) \ll 1$) is assumed such that $\overline{\mathrm{ISR}}_{ij} \approx \mathrm{var}(G_{ij})$. In this section, we do not use the ISR matrix but instead directly derive the iCRB for $\mathbf{G}$.

with

$$\mathbf{R}_{\vartheta} = (\mathbf{M}_1 - \mathbf{M}_2\mathbf{M}_1^{-*}\mathbf{M}_2^*)^*. \tag{3.21}$$

As shown in [35], $\mathbf{R}_{\vartheta}$ can be calculated as

$$\mathbf{R}_{\vartheta} = \sum_{i=1}^{N} d_i \mathbf{L}_{ii} \otimes \mathbf{L}_{ii} + \sum_{i=1}^{N} \sum_{j=1,j\neq i}^{N} a_{ij}\mathbf{L}_{ii} \otimes \mathbf{L}_{jj} + \sum_{i=1}^{N} \sum_{j=1,j\neq i}^{N} b_{ij}\mathbf{L}_{ij} \otimes \mathbf{L}_{ji} \tag{3.22}$$

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 (3.22) denotes an $N \times N$ matrix with a 1 at the (i, j) position and 0's elsewhere.

The parameters $\eta_i, \kappa_i, \beta_i, \xi_i$ and γ_j are defined as

$$\eta_i = \mathrm{E}\left[|s_i|^2|\varphi_i(s_i)|^2\right] > 1, \tag{3.23}$$

$$\kappa_i = \mathrm{E}\left[|\varphi_i(s_i)|^2\right] \geq 1, \tag{3.24}$$

$$\beta_i = \mathrm{E}\left[s_i^2(\varphi_i^*(s_i))^2\right] \in \mathbb{C}, \tag{3.25}$$

$$\xi_i = \mathrm{E}\left[(\varphi_i^*(s_i))^2\right] \in \mathbb{C}, \tag{3.26}$$

$$\gamma_j = \mathrm{E}\left[s_j^2\right] \in \mathbb{R}. \tag{3.27}$$

Properties and other equivalent forms of these parameters can be found in the appendix of [35].

$\mathbf{R}_{\vartheta}$ has a special sparse structure which is illustrated below for $N = 3$:

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

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

$$
\begin{aligned}
\mathbf{R}_{\vartheta}^{-1} &= \sum_{i=1}^{N} \frac{1}{d_i} \mathbf{L}_{ii} \otimes \mathbf{L}_{ii} + \sum_{i=1}^{N} \sum_{\substack{j=1 \\ j \neq i}}^{N} \frac{a_{ji}}{a_{ij}a_{ji} - b_{ij}b_{ji}} \mathbf{L}_{ii} \otimes \mathbf{L}_{jj} \\
&\quad + \sum_{i=1}^{N} \sum_{\substack{j=1 \\ j \neq i}}^{N} \frac{-b_{ij}}{a_{ij}a_{ji} - b_{ij}b_{ji}} \mathbf{L}_{ij} \otimes \mathbf{L}_{ji} \\
&= \sum_{i=1}^{N} f_i \mathbf{L}_{ii} \otimes \mathbf{L}_{ii} + \sum_{i=1}^{N} \sum_{\substack{j=1 \\ j \neq i}}^{N} g_{ij} \mathbf{L}_{ii} \otimes \mathbf{L}_{jj} + \sum_{i=1}^{N} \sum_{\substack{j=1 \\ j \neq i}}^{N} h_{ij} \mathbf{L}_{ij} \otimes \mathbf{L}_{ji} \quad (3.28)
\end{aligned}
$$

with

$$
f_i = \frac{1}{d_i} = \frac{\eta_i - 1}{(\eta_i - 1)^2 - |\beta_i - 1|^2}, \tag{3.29}
$$

$$
g_{ij} = \frac{a_{ji}}{a_{ij}a_{ji} - b_{ij}b_{ji}} = \frac{\kappa_j(\kappa_i \kappa_j - 1) - |\gamma_i \xi_j|^2 \kappa_i}{(\kappa_i \kappa_j - 1)^2 + |\gamma_i \gamma_j \xi_i \xi_j - 1|^2 - 1 - \kappa_i^2 |\gamma_i \xi_j|^2 - \kappa_j^2 |\gamma_j \xi_i|^2}, \tag{3.30}
$$

$$
h_{ij} = \frac{-b_{ij}}{a_{ij}a_{ji} - b_{ij}b_{ji}} = \frac{\gamma_j^* \xi_i^* \kappa_j + \gamma_i \xi_j \kappa_i}{(\kappa_i \kappa_j - 1)^2 + |\gamma_i \gamma_j \xi_i \xi_j - 1|^2 - 1 - \kappa_i^2 |\gamma_i \xi_j|^2 - \kappa_j^2 |\gamma_j \xi_i|^2}. \tag{3.31}
$$

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 $L^{-1}\mathbf{R}_{\vartheta}^{-1}$:

$$
\mathrm{var}(\hat{G}_{ii}) \geq \frac{1}{L} f_i = \frac{1}{L} \frac{\eta_i - 1}{(\eta_i - 1)^2 - |\beta_i - 1|^2}, \tag{3.32}
$$

$$
\mathrm{var}(\hat{G}_{ij}) \geq \frac{1}{L} g_{ij} = \frac{1}{L} \frac{\kappa_j(\kappa_i \kappa_j - 1) - |\gamma_i \xi_j|^2 \kappa_i}{(\kappa_i \kappa_j - 1)^2 + |\gamma_i \gamma_j \xi_i \xi_j - 1|^2 - 1 - \kappa_i^2 |\gamma_i \xi_j|^2 - \kappa_j^2 |\gamma_j \xi_i|^2}. \tag{3.33}
$$

Note that $L^{-1}\mathbf{R}_{\vartheta}^{-1}$ is the iCRB for ϑ as in (3.9). In order to get the complete iCRB for $\begin{bmatrix} \vartheta \\ \vartheta^* \end{bmatrix}$ as in (3.8), we would also need $\mathbf{P}_{\vartheta} = -\mathbf{R}_{\vartheta}^{-1}\mathbf{Q}_{\vartheta} = -\mathbf{R}_{\vartheta}^{-1}\mathbf{M}_2^*\mathbf{M}_1^{-1}$.

It can be shown in a similar way

$$\mathbf{P}_{\boldsymbol{\vartheta}} = \sum_{i=1}^{N} \tilde{f}_i \mathbf{L}_{ii} \otimes \mathbf{L}_{ii} + \sum_{i=1}^{N} \sum_{\substack{j=1 \\ j \neq i}}^{N} \left(\tilde{g}_{ij} \mathbf{L}_{ii} \otimes \mathbf{L}_{jj} + \tilde{h}_{ij} \mathbf{L}_{ij} \otimes \mathbf{L}_{ji} \right) \qquad (3.34)$$

has the same form as $\mathbf{R}_{\boldsymbol{\vartheta}}^{-1}$ in (3.28) with

$$\tilde{f}_i = -\frac{f_i(\beta_i - 1)^*}{\eta_i - 1} = \frac{-(\beta_i - 1)^*}{(\eta_i - 1)^2 - |\beta_i - 1|^2}, \qquad (3.35)$$

$$\tilde{g}_{ij} = -\frac{g_{ij}\gamma_j^*\xi_i^* + h_{ij}}{\kappa_i} = \frac{-(\kappa_j^2 - |\gamma_i\xi_j|^2)\gamma_j^*\xi_i^*\gamma_i\xi_j}{(\kappa_i\kappa_j - 1)^2 + |\gamma_i\gamma_j\xi_i\xi_j - 1|^2 - 1 - \kappa_i^2|\gamma_i\xi_j|^2 - \kappa_j^2|\gamma_j\xi_i|^2}, \qquad (3.36)$$

$$\tilde{h}_{ij} = -\frac{g_{ij} + \gamma_i^*\xi_i^* h_{ij}}{\kappa_j} = \frac{1 - \kappa_i\kappa_j - (\gamma_j\xi_i\gamma_i\xi_j)^*}{(\kappa_i\kappa_j - 1)^2 + |\gamma_i\gamma_j\xi_i\xi_j - 1|^2 - 1 - \kappa_i^2|\gamma_i\xi_j|^2 - \kappa_j^2|\gamma_j\xi_i|^2}. \qquad (3.37)$$

Note that according to (3.28) and (3.34) the iCRB for $\mathbf{G} = \mathbf{WA}$ 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 any other sources. Note that (3.32) and (3.33) cannot be used as a bound for real ICA since the FIM would be singular.

3.2.2.2 CRB for the Demixing Matrix W

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

$$\text{cov}(\hat{\boldsymbol{\theta}}) \geq L^{-1}\mathbf{R}_{\boldsymbol{\theta}}^{-1} = L^{-1}(\mathbf{I} \otimes \mathbf{W}^T)\mathbf{R}_{\boldsymbol{\vartheta}}^{-1}(\mathbf{I} \otimes \mathbf{W}^*) \qquad (3.38)$$

holds for any unbiased estimator $\hat{\boldsymbol{\theta}}$ for $\boldsymbol{\theta}$. See [35] for a more detailed expression of the CRB for $\mathbf{W}$.

3.2.3 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.

3.2.3.1 Case A: All Sources Are Circular Complex

If all sources are circular complex, $\gamma_i = 0$ and $\beta_i = \eta_i$ [35]. 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 [24] for G_{ii}: As noted at the end of Sect. 3.2.2.1, G_{ii} is decoupled from G_{ij} and G_{jj} and hence it is sufficient to consider the constrained CRB for G_{ii} alone.

The constrained CRB for G_{ii} is given by [35]

$$\text{var}(\hat{G}_{ii}) \geq \frac{1}{4L(\eta_i - 1)}. \tag{3.39}$$

The bound in (3.39) is valid for a phase-constrained G_{ii} such that $G_{ii} \in \mathbb{R}$. Equation (3.39) looks similar to the real case (3.90) except for a factor of 4 since η_i is defined using Wirtinger derivatives instead of real derivatives.

For $\text{var}(\hat{G}_{ij})$ we get from (3.33)

$$\text{var}(\hat{G}_{ij}) \geq \frac{1}{L} \frac{\kappa_j}{\kappa_i \kappa_j - 1}, \tag{3.40}$$

which again looks the same as in the real case (3.91). 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 $\text{var}(\hat{G}_{ij}) \to \infty$. This corresponds to the known fact that circular complex Gaussian sources cannot be separated by ICA.

3.2.3.2 Case 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 in (3.86)–(3.89) with $c = 1$ that

$$\text{var}(\hat{G}_{ii}) \geq \frac{1}{L} \frac{1}{4\gamma_i^2}, \tag{3.41}$$

$$\text{var}(\hat{G}_{ij}) \geq \frac{1}{L} \frac{\gamma_i^2 + \gamma_j^2 - 2\gamma_i^2 \gamma_j^2}{(\gamma_j^2 - \gamma_i^2)^2} (1 - \gamma_i^2)$$

$$= \frac{1 - \gamma_i^2}{2L} \left[\frac{1 - \gamma_i \gamma_j}{(\gamma_i - \gamma_j)^2} + \frac{1 + \gamma_i \gamma_j}{(\gamma_i + \gamma_j)^2} \right]. \tag{3.42}$$

Note that (3.42) is exactly the same result as obtained in [48] 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 [19, 29] which showed that ICA fails for two or more noncircular Gaussian signals with same γ_i.

3.2.4 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 $E[|s|^2] = 1$ and noncircularity index $\gamma \in [0, 1]$ can be written as [40]

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

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 1. 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 consider the following different ICA estimators[7]: Complex ML-ICA [29], complex ICA by entropy bound minimization (ICA-EBM) [30], noncircular complex ncFastICA (ncFastICA) [39], adaptable complex maximization of non-Gaussianity (ACMN) [38] and strong uncorrelating transform (SUT) [18, 44]. The properties and assumptions of the five different ICA algorithms are summarized in Table 3.1.

We want to compare the separation performance of ICA with respect to the iCRB and hence we define the performance metric as in [45]: 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:

$$\text{SIR} = \frac{1}{N}\sum_{i=1}^{N}\frac{E\left[|G_{ii}|^2\right]}{\sum_{j\neq i}E\left[|G_{ij}|^2\right]} = \frac{1}{N}\sum_{i=1}^{N}\frac{1 + \text{var}(G_{ii})}{\sum_{j\neq i}\text{var}(G_{ij})}. \tag{3.43}$$

---

[7] Note that many alternative ICA estimators such as [7, 10, 14, 17, 20] exist.

Table 3.1 Considered separation algorithms and their properties

Algorithm	Pdf model/φ	Separation principle
Complex ML-ICA (ML-ICA)	True pdf of the sources	Non-Gaussianity and noncircularity
Complex ICA by Entropy Bound Minimization (ICA-EBM)	Adaptive	Non-Gaussianity and noncircularity
Noncircular Complex FastICA (ncFastICA)	Fixed φ	Non-Gaussianity
Adaptable Complex Maximization of Non-Gaussianity (ACMN)	Adaptive φ	Non-Gaussianity
Strong Uncorrelating Transform (SUT)	–	noncircularity

In (3.43), 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.

3.2.4.1 Case 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. Figure 3.1 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, all ICA algorithms work quite well except for $c \approx 1$ (Gaussian). ML-ICA (Fig. 3.1b) achieves the best performance followed by ICA-EBM (Fig. 3.1c) and ACMN (Fig. 3.1f). ncFastICA with kurtosis cost function achieves better performance for sub-Gaussian sources ($c > 1$) than for super-Gaussian sources ($c < 1$), whereas ncFastICA with square root (sqrt) nonlinearity works better for super-Gaussian sources than for sub-Gaussian sources. However, as also mentioned in [30], the square root nonlinearity leads overall to the best performance and hence we only consider ncFastICA with this nonlinearity in the following. As expected, SUT fails since it only uses noncircularity for separation and hence we do not show the results. The reason why ML-ICA outperforms ICA-EBM is 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.
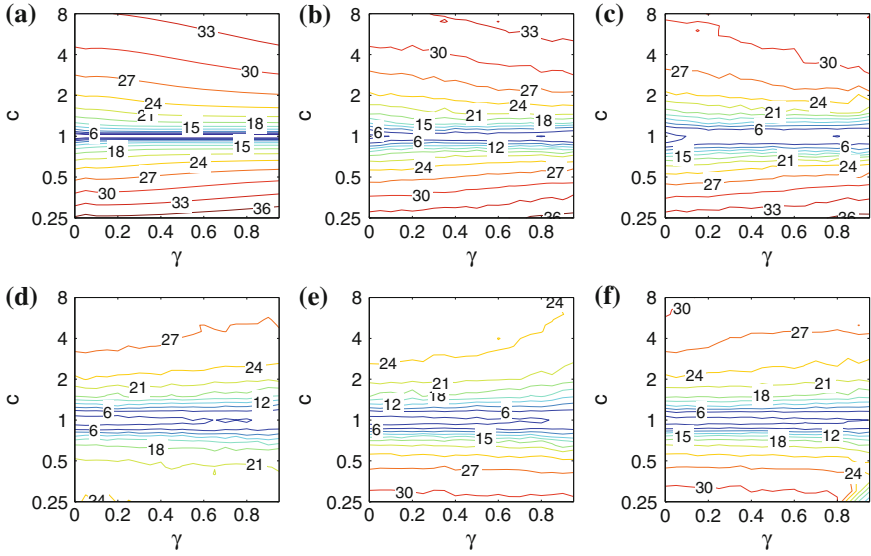
Fig. 3.1 Comparison of signal-to-interference ratio [dB] of different ICA estimators with CRB, sample size $L = 1000$, all sources follow a generalized Gaussian distribution with $c_i = c$ and $\gamma_i = \gamma$. iCRB (**a**), ML-ICA (**b**), ICA-EBM (**c**), ncFastICA kurtosis (**d**), ncFastICA sqrt (**e**), ACMN (**f**)

3.2.4.2 Case B: All Sources Have Different Shape Parameters and Different Noncircularities

Now we study the performance when the sources follow a GGD with different shape parameters $c_1 = 1, c_2 = c, c_3 = 1/c$ and different noncircularity indices $\gamma_i = (i - 1)\Delta\gamma$. Figure 3.2 shows that the SIR given by the iCRB increases both with increasing non-Gaussianity of source 2 and 3 (i.e., $c < 1$) as well as increasing difference in noncircularity indices $\Delta\gamma$. ML-ICA achieves again the best performance, followed by 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 $p_{s_i}(s_i)$ whereas the nonlinearities used in ICA-EBM are fixed a priori. Although ncFastICA and ACMN exploit the noncircularity of the sources to improve the convergence, their cost function only uses non-Gaussianity and not noncircularity. This is reflected clearly in Fig. 3.2 since performance for ncFastICA and ACMN is almost constant for different $\Delta\gamma$. SUT uses only noncircularity for separation, and hence performance is almost constant for different c. SUT can work quite well, as long as $\Delta\gamma$ is large enough. Only ML-ICA and ICA-EBM use both non-Gaussianity and noncircularity and hence the contour lines in Fig. 3.2b, c resemble those of the CRB Fig. 3.2a.
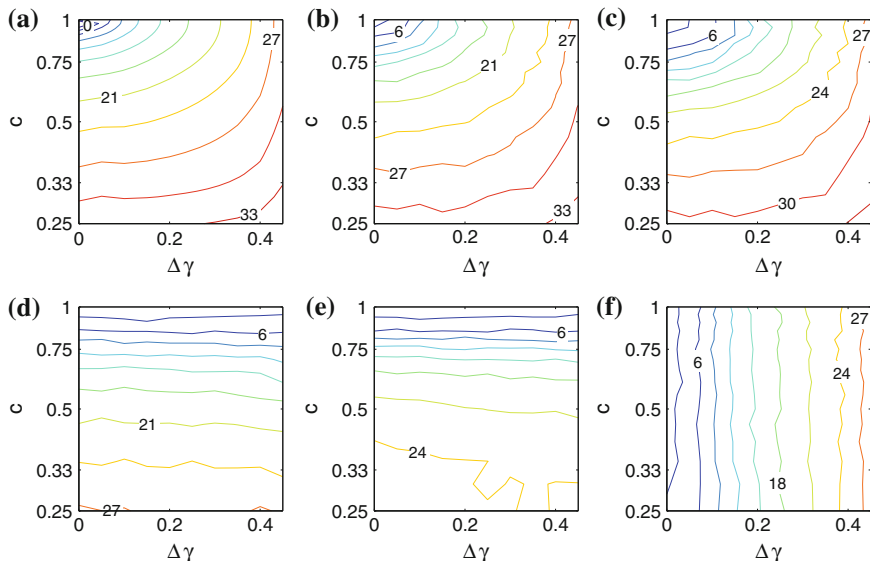
Fig. 3.2 Comparison of signal-to-interference ratio [dB] of different ICA estimators with iCRB, sample size $L = 1000$, all sources follow a generalized Gaussian distribution with $c_1 = 1, c_2 = c, c_3 = 1/c, \gamma_i = (i - 1)\Delta\gamma$. iCRB (**a**), ML-ICA (**b**), ICA-EBM (**c**), ncFastICA sqrt (**d**), ACMN (**e**), SUT (**f**)

3.2.4.3 Performance as a Function of the Sample Size

Here, we study the performance as a function of sample size L. Clearly, Fig. 3.3 shows that for circular *non-Gaussian* sources and limited sample size L, ML-ICA achieves the best performance followed by ACMN and then ICA-EBM. The reason why ACMN outperforms ICA-EBM for circular sources could be the fact that ACMN needs to adapt less parameters since it uses only non-Gaussianity. As expected, SUT fails since it only uses noncircularity for separation. For circular super-Gaussian sources (Fig. 3.3a), ACMN and ncFastICA perform almost the same. For sub-Gaussian sources (Fig. 3.3b), the sqrt nonlinearity is sub-optimal as shown in the larger error of ncFastICA. Figure 3.4a shows results for noncircular *Gaussian sources* with distinct noncircularity indices: SUT and ML-ICA perform equally well since for noncircular Gaussian sources they are equivalent and asymptotically optimal. ICA-EBM approaches the performance of SUT and ML-ICA for large enough sample size. ncFastICA and ACMN which use only non-Gaussianity for separation fail. Figure 3.4b, c shows results for noncircular super-Gaussian ($c = 0.5$) and sub-Gaussian ($c = 6$) sources with distint noncircularity indices: With limited sample size, ML-ICA achieves again the best performance followed by ICA-EBM. For a large sample size ($L \geq 1000$) and a wide range of distributions including strongly super-Gaussian but excluding strongly sub-Gaussian sources, ICA-EBM comes close to the performance of ML-ICA, see Figs. 3.1, 3.3, 3.4. The reason for
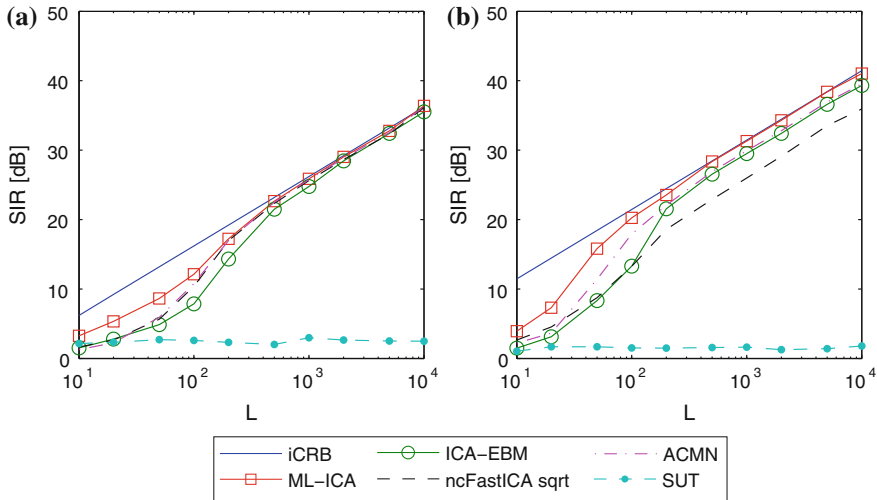
Fig. 3.3 Performance as a function of sample size L, circular GGD sources. $c = 0.5$ (**a**), $c = 6$ (**b**)

this behavior is that ML-ICA uses nonlinearities matched to the source distributions while ICA-EBM uses a linear combination of prespecified nonlinear functions. These could be extended to improve performance for strongly sub-Gaussian sources. The performance of ncFastICA and ACMN is quite far from that given by the iCRB since these two algorithms do not use noncircularity for separation. For signals with distinct noncircularity indices, SUT can achieve decent separation, but for strongly non-Gaussian signals the performance is quite far from that given by the iCRB (see also Fig. 3.2).

3.2.5 Conclusion

In this section, 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 five ICA algorithms have shown that for sources following a noncircular complex generalized Gaussian distribution, some algorithms can achieve a signal-to-interference ratio (SIR) close to that of the CRB. Among the studied algorithms, complex ML-ICA and ICA-EBM perform best. The complex ML-ICA algorithm, which uses for each source a nonlinearity matched to its pdf, outperforms ICA-EBM especially for small
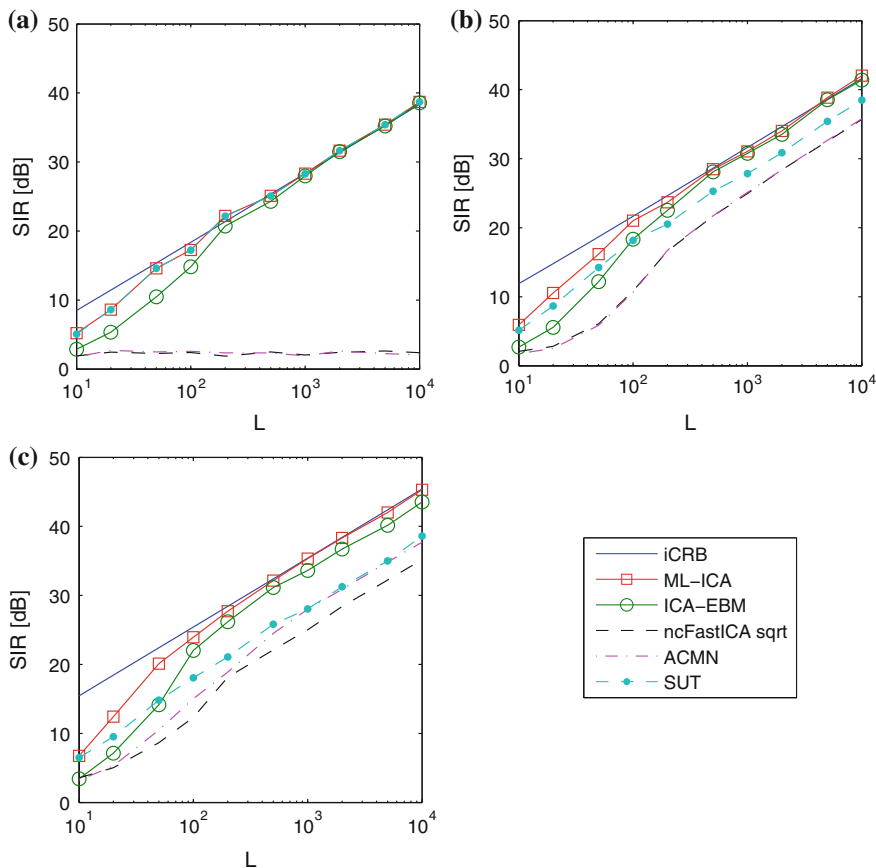
Fig. 3.4 Performance as a function of sample size L, noncircular GGD sources with $c_i = c$ and $\gamma_i = (i-1)\Delta\gamma$. $c = 1$, $\Delta\gamma = 0.45$ (**a**), $c = 0.5$, $\Delta\gamma = 0.45$ (**b**), $c = 6$, $\Delta\gamma = 0.45$ (**c**)

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.

3.3 Solution of Linear Complex ICA in the Presence of Noise

In this section, we study the bias of the demixing matrix in linear noisy ICA from the inverse mixing matrix. We first derive the ICA solution for the general complex determined case. We then show how the circular complex case and the real case can be derived as special cases. Next, we verify the results using simulations. Finally, we extend our derivations to the overdetermined case with circular complex noise.

3.3.1 Signal Model and Assumptions

We assume the linear noisy mixing model

$$\mathbf{x} = \mathbf{As} + \mathbf{v} \tag{3.44}$$

where $\mathbf{x} \in \mathbb{C}^N$ are N linear combinations of N original signals $\mathbf{s} \in \mathbb{C}^N$ with additive noise $\mathbf{v} \in \mathbb{C}^N$. Here, all signals are modeled as temporally iid. In addition to the assumptions A1 and A2 (invertibility of mixing matrix $\mathbf{A}$ and assumptions about the source signals $\mathbf{s}$) defined in Sect. 3.2.1, we make the following two assumptions regarding the noise $\mathbf{v}$:

1. $\mathbf{v} = [v_1, \ldots, v_N]^T \in \mathbb{C}^N$ are N random variables with zero mean and the covariance matrix $E[\mathbf{vv}^H] = \sigma^2 \mathbf{R_v}$. $\sigma^2 = \frac{1}{N} \text{tr}\left[E(\mathbf{vv}^H)\right]$ is the average variance of $\mathbf{v}$ and $\text{tr}(\mathbf{R_v}) = N$. $\bar{\mathbf{R}}_\mathbf{v} = \frac{1}{\sigma^2} E[\mathbf{vv}^T]$ is the normalized pseudo-covariance matrix. $\bar{\mathbf{R}}_\mathbf{v} = \mathbf{0}$ if $\mathbf{v}$ is circular complex. The pdf of $\mathbf{v}$ is arbitrary but assumed to be symmetric, i.e., $p_\mathbf{v}(\mathbf{v}) = p_\mathbf{v}(-\mathbf{v})$. This implies $E(\prod_{i=1}^{N} v_i^{k_i} (v_i^*)^{\tilde{k}_i}) = 0$ for $\sum_{i=1}^{N} \left(k_i + \tilde{k}_i\right)$ odd.
2. $\mathbf{s}$ and $\mathbf{v}$ are independent.

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

$$\mathbf{y} = \mathbf{Wx} = \mathbf{WAs} + \mathbf{Wv} \tag{3.45}$$

so that $\mathbf{y}$ is "as close to $\mathbf{s}$" as possible according to some metric.

3.3.2 KLD-Based ICA for Determined Case

We focus on the ICA solution based on the KLD

$$D_{\text{KL}}(\mathbf{W}) = \int p_\mathbf{y}(\mathbf{y}; \mathbf{W}) \ln \frac{p_\mathbf{y}(\mathbf{y}; \mathbf{W})}{p_\mathbf{s}(\mathbf{y})} d\mathbf{y} \tag{3.46}$$

where $p_\mathbf{y}(\mathbf{y}; \mathbf{W})$ is the pdf of $\mathbf{y}$. It depends on the pdf of observation $\mathbf{x}$, i.e., on the pdf of the original source signals $\mathbf{s}$ and noise $\mathbf{v}$, as well as on the demixing matrix $\mathbf{W}$. $p_\mathbf{s}(\mathbf{s}) = \prod_{i=1}^{N} p_{s_i}(s_i)$ is the assumed pdf of the original signals. We assume that we have perfect knowledge about the distribution of the original signals and $p_\mathbf{s}(\mathbf{s})$ is identical to the true pdf $p_\mathbf{s}^0(\mathbf{s})$ of $\mathbf{s}$. The KLD is known to have the following properties:

- $D_{\text{KL}}(\mathbf{W}) \geq 0$ for any $p_\mathbf{y}(\mathbf{y}; \mathbf{W})$ and $p_\mathbf{s}(\mathbf{y})$.
- $D_{\text{KL}}(\mathbf{W}) = 0$ iff $p_\mathbf{y}(\mathbf{y}; \mathbf{W}) = p_\mathbf{s}(\mathbf{y})$.

This means, minimizing the KLD with respect to $\mathbf{W}$ is equivalent to making the pdf of the demixed signals $\mathbf{y}$ as similar as possible to the pdf of the source signals $p_\mathbf{s}(\mathbf{s})$.

Since we assume $p_\mathbf{s}(\mathbf{s}) = \prod_{i=1}^{N} p_{s_i}(s_i)$, minimizing KLD corresponds to making (a) y_i as independent as possible and (b) y_i to have a pdf as close as possible to $p_{s_i}(s_i)$. This has been stated as "total mismatch = deviation from independence + marginal mismatch" by Cardoso in [9]. The ICA solution $\mathbf{W}_{\text{ICA}}$ for the demixing matrix based on KLD is given by

$$\mathbf{W}_{\text{ICA}} = \arg \min_{\mathbf{W}} D_{\text{KL}}(\mathbf{W}). \tag{3.47}$$

In the following, we will first derive the ICA solution for the general noncircular complex case. The circular complex case and the real case are discussed as two special cases.

3.3.2.1 General Noncircular Complex Case

The KLD cost function of a complex demixing matrix $\mathbf{W}$ is a function of the real and imaginary part of $\mathbf{W}$. Using the Wirtinger calculus (see [21, 44] and the summary in Sect. 3.1.1.2), we can also write it as a function of $\mathbf{W}$ and $\mathbf{W}^*$:

$$D_{\text{KL}}(\mathbf{W}, \mathbf{W}^*) = \int p_\mathbf{y}(\mathbf{y}, \mathbf{y}^*; \mathbf{W}, \mathbf{W}^*) \ln \frac{p_\mathbf{y}(\mathbf{y}, \mathbf{y}^*; \mathbf{W}, \mathbf{W}^*)}{p_\mathbf{s}(\mathbf{y}, \mathbf{y}^*)} d\mathbf{y}. \tag{3.48}$$

The derivative $\frac{\partial D_{\text{KL}}(\mathbf{W}, \mathbf{W}^*)}{\partial \mathbf{W}^*}$ of the KLD cost function in (3.48) is given by [21]

$$\frac{\partial D_{\text{KL}}(\mathbf{W}, \mathbf{W}^*)}{\partial \mathbf{W}^*} = -\mathbf{W}^{-H} + \text{E}\left[\boldsymbol{\varphi}(\mathbf{y}, \mathbf{y}^*)\mathbf{x}^H\right], \tag{3.49}$$

where $\boldsymbol{\varphi}(\mathbf{y}, \mathbf{y}^*) = [\varphi_1(y_1, y_1^*), \ldots, \varphi_N(y_N, y_N^*)]^T$ and $\varphi_i(s_i, s_i^*) = -\frac{\partial \ln p_{s_i}(s_i, s_i^*)}{\partial s_i^*}$. The derivative $\frac{\partial}{\partial s^*}$ is also defined using the Wirtinger calculus.

A necessary condition for minimizing $D_{\text{KL}}(\mathbf{W}, \mathbf{W}^*)$ at $\mathbf{W} = \mathbf{W}_{\text{ICA}}$ is

$$\left.\frac{\partial D_{\text{KL}}(\mathbf{W}, \mathbf{W}^*)}{\partial \mathbf{W}^*}\right|_{\mathbf{W}=\mathbf{W}_{\text{ICA}}} \overset{!}{=} \mathbf{0} \quad \text{or} \quad \text{E}(\boldsymbol{\varphi}(\mathbf{y}_{\text{ICA}}, \mathbf{y}_{\text{ICA}}^*)\mathbf{y}_{\text{ICA}}^H) \overset{!}{=} \mathbf{I} \tag{3.50}$$

with $\mathbf{y}_{\text{ICA}} = \mathbf{W}_{\text{ICA}}\mathbf{x} = \mathbf{W}_{\text{ICA}}\mathbf{A}\mathbf{s} + \mathbf{W}_{\text{ICA}}\mathbf{v} = \hat{\mathbf{y}} + \mathbf{W}_{\text{ICA}}\mathbf{v}$. An equivalent condition to $\text{E}(\boldsymbol{\varphi}(\mathbf{y}_{\text{ICA}}, \mathbf{y}_{\text{ICA}}^*)\mathbf{y}_{\text{ICA}}^H) \overset{!}{=} \mathbf{I}$ in (3.50) is

$$\text{E}(\boldsymbol{\varphi}(\mathbf{y}_{\text{ICA}}, \mathbf{y}_{\text{ICA}}^*)\mathbf{y}_{\text{ICA}}^H)^* = \text{E}(\boldsymbol{\varphi}^*(\mathbf{y}_{\text{ICA}}, \mathbf{y}_{\text{ICA}}^*)\mathbf{y}_{\text{ICA}}^T) \overset{!}{=} \mathbf{I} \tag{3.51}$$

which we will use in the following to facilitate comparison with Sect. 3.2. The properties of the ICA solution based on KLD are:

- $\mathbf{W}_{\text{ICA}} = \mathbf{A}^{-1}$ if $\sigma^2 = 0$ (no noise) and $p_\mathbf{s}(\mathbf{s}) = p_\mathbf{s}^0(\mathbf{s})$.
- To compute $\mathbf{W}_{\text{ICA}}$, we do not need to know $\mathbf{A}$ or $\mathbf{s}$, but the pdf $p_\mathbf{s}(\mathbf{s}) = \prod_{i=1}^{N} p_{s_i}(s_i)$ is required. All $p_{s_i}(s_i)$ must either be non-Gaussian or Gaussian with distinct noncircularity indices.

- No permutation ambiguity if $p_{s_i}(\cdot) \neq p_{s_j}(\cdot) \ \forall i \neq j$.
- There is no scaling ambiguity if $p_{s_i}(s_i) = p_i^0(s_i)$ is known $\forall i$. Only a phase ambiguity remains if $p_{s_i}(s_i)$ is circular.

As shown in Appendix 2, the ICA solution for the general noncircular complex case can be derived approximately using a two-step perturbation analysis for low noise and is given by

$$\mathbf{W}_{\text{ICA}} = (\mathbf{I} + \sigma^2 \mathbf{C})\mathbf{A}^{-1} + \mathcal{O}(\sigma^4). \tag{3.52}$$

The elements of $\mathbf{C}$ can be obtained from (3.97) and (3.98). If $p_{\mathbf{s}}(\mathbf{s})$ is symmetric in the real or imaginary part of $\mathbf{s}$, they are given by (3.99) and (3.100).

For comparison, we consider the linear MMSE estimator

$$\mathbf{W}_{\text{MMSE}} = \mathbf{A}^H \left(\mathbf{A}\mathbf{A}^H + \sigma^2 \mathbf{R}_{\mathbf{v}} \right)^{-1} \tag{3.53}$$

$$= \left[\mathbf{I} - \sigma^2 \mathbf{R}_{-1} \right] \mathbf{A}^{-1} + \mathcal{O}(\sigma^4). \tag{3.54}$$

where the last line is a first-order Taylor series expansion in σ^2 and $\mathbf{R}_{-1} = \mathbf{A}^{-1} \mathbf{R}_{\mathbf{v}} \mathbf{A}^{-H}$. Comparing (3.54) with (3.52) we see that $\mathbf{W}_{\text{ICA}}$ and $\mathbf{W}_{\text{MMSE}}$ are similar if $\mathbf{C} \approx -\mathbf{R}_{-1}$.

3.3.2.2 Circular Complex Case

We assume now that the source signals $\mathbf{s}$ and the noise $\mathbf{v}$ are circular. Hence, both the noncircularity index of the sources γ and the pseudo-covariance matrix $\bar{\mathbf{R}}_{\mathbf{v}}$ are zero. As a consequence, (3.99) and (3.100) simplify to

$$C_{ii} = -\frac{\kappa_i + \lambda_i}{1 + \rho_i + \delta_i} \left[\mathbf{R}_{-1} \right]_{ii} \in \mathbb{R},$$

$$C_{ij} = -\frac{\kappa_j (\kappa_i - 1)}{\kappa_i \kappa_j - 1} \left[\mathbf{R}_{-1} \right]_{ij} \in \mathbb{C} \quad (i \neq j). \tag{3.55}$$

3.3.2.3 Real Case

For real signals and noise, we have

$$\gamma_i = 1, \quad \mathbf{R}_{\mathbf{v}} = \bar{\mathbf{R}}_{\mathbf{v}}. \tag{3.56}$$

In the derivation of $\mathbf{W}_{\text{ICA}}$ we have considered Taylor series expansions of $\boldsymbol{\varphi}(\mathbf{y})$ using Wirtinger derivatives. The Wirtinger derivatives $\partial/\partial s$ and $\partial/\partial s^*$ of $\varphi(s) \in \mathbb{R}$ are now identical (see (3.4)) and hence

$$\xi_i = \kappa_i, \quad \rho_i = \delta_i, \quad \lambda_i = \omega_i = \tau_i. \tag{3.57}$$

Furthermore, the Wirtinger derivatives of $\varphi(s) \in \mathbb{R}$ are identical to the real derivatives except for a factor of $\frac{1}{2}$ (see (3.4)). Hence it holds

$$\kappa_i = \frac{\mathring{\kappa}_i}{2}, \quad \rho_i = \frac{\mathring{\rho}_i}{2}, \quad \lambda_i = \frac{\mathring{\lambda}_i}{4}, \tag{3.58}$$

where $\mathring{\kappa}_i$, $\mathring{\rho}_i$ and $\mathring{\lambda}_i$ are defined using real derivatives of $\varphi(s)$, denoted by $(\cdot)'$:

$$\mathring{\kappa}_i = \mathrm{E}(\varphi_i'(s_i)) = \int \frac{d}{ds_i} \left(\frac{-p_{s_i}'(s_i)}{p_{s_i}(s_i)} \right) p_i^0(s_i) ds_i,$$

$$\mathring{\rho}_i = \mathrm{E}(\varphi_i'(s_i)s_i^2) = \int \frac{d}{ds_i} \left(\frac{-p_{s_i}'(s_i)}{p_{s_i}(s_i)} \right) s_i^2 p_i^0(s_i) ds_i,$$

$$\mathring{\lambda}_i = \mathrm{E}(\varphi_i''(s_i)s_i) = \int \frac{d^2}{ds_i^2} \left(\frac{-p_{s_i}'(s_i)}{p_{s_i}(s_i)} \right) s_i p_i^0(s_i) ds_i. \tag{3.59}$$

Using (3.56)–(3.58), we get from (3.99) and (3.100)

$$C_{ii} = -\frac{\mathring{\kappa}_i + \frac{1}{2}\mathring{\lambda}_i}{1 + \mathring{\rho}_i} \left[\mathbf{R}_{-1} \right]_{ii} = -M_{ii} \left[\mathbf{R}_{-1} \right]_{ii},$$

$$C_{ij} = -\frac{\mathring{\kappa}_j(\mathring{\kappa}_i - 1)}{\mathring{\kappa}_i \mathring{\kappa}_j - 1} \left[\mathbf{R}_{-1} \right]_{ij} = -M_{ij} \left[\mathbf{R}_{-1} \right]_{ij} \quad (i \neq j). \tag{3.60}$$

where $M_{ii} = \frac{\mathring{\kappa}_i + \mathring{\lambda}_i/2}{1 + \mathring{\rho}_i}$ and $M_{ij} = \frac{\mathring{\kappa}_j(\mathring{\kappa}_i - 1)}{\mathring{\kappa}_i \mathring{\kappa}_j - 1}$. Note that (3.60) corresponds to the results in [32].

3.3.3 Results for Complex Generalized Gaussian Distribution

We study KLD-ICA for $N = 3$ sources with spatially white Gaussian noise with $E[\mathbf{v}\mathbf{v}^H] = \sigma^2 \mathbf{I}$ and the square mixing matrix $\mathbf{A} = [a_{mn}]$, where $a_{mn} = e^{-j\pi m \sin \theta_n}$ and $\theta_n = -60°, 0°, 60°$. As proposed in [26], we use the signal-to-interference-plus-noise ratio (SINR) to evaluate separation performance. For spatially uncorrelated noise, we compute the SINR for a given demixing matrix $\mathbf{W}$ by averaging the SINR for each source i

$$\mathrm{SINR} = \frac{1}{N} \sum_{i=1}^{N} \frac{|[\mathbf{WA}]_{ii}|^2}{\sum_{j \neq i} |[\mathbf{WA}]_{ij}|^2 + \sigma^2 \sum_j |W_{ij}|^2}. \tag{3.61}$$

The term $|[\mathbf{WA}]_{ii}|^2$ reflects the power of the desired source i in the demixed signal y_i. The term $\sum_{j \neq i} |[\mathbf{WA}]_{ij}|^2$ corresponds to the power of the interfering signals

$j \neq i$ in the demixed signal y_i and $\sigma^2 \sum_j |\mathbf{W}_{ij}|^2$ is the noise power in the demixed signal y_i. For the remainder of this section, the signal-to-noise ratio (SNR) is defined as that before the mixing process and not at the sensors, i.e., SNR $= \frac{\mathrm{E}[s^2]}{\sigma^2} = \frac{1}{\sigma^2}$. It can be shown that among all linear demixing matrices $\mathbf{W}$, $\mathbf{W}_{\mathrm{MMSE}}$ from (3.53) is the one which maximizes the SINR [28]. We compare the SINR of the theoretical ICA solution $\mathbf{W}_{\mathrm{ICA}}$ from (3.52), the average SINR of $\hat{\mathbf{W}}_{\mathrm{ICA}}$ obtained from 100 runs of KLD-based ICA using L samples and the SINR of $\mathbf{W}_{\mathrm{MMSE}}$ from (3.53). The ICA algorithm is initialized with $\mathbf{W} = \mathbf{I}$ and performs gradient descent using the relative gradient [12], i.e., postmultiplies the gradient of KLD (3.49) by $\mathbf{W}^H\mathbf{W}$. We normalize each row of the relative gradient, resulting in an adaptive step size for each source. In the derivation of the theoretical solution $\mathbf{W}_{\mathrm{ICA}}$, we evaluated all expectations exactly. Hence $\mathbf{W}_{\mathrm{ICA}}$ only accounts for the bias from $\mathbf{A}^{-1}$ but not for estimation variance whereas $\hat{\mathbf{W}}_{\mathrm{ICA}}$ contains both factors.

In the following, all sources are GGD with the same shape parameter $c_i = c$. The noncircular complex GGD with zero mean and $E[|s|^2] = 1$ has already been introduced in Sect. 3.2.4. By integration in polar coordinates, it can be shown that

$$\kappa = \int \frac{\partial \varphi^*}{\partial s^*} p_{\mathbf{s}}^0(s) ds = \frac{c^2 \Gamma(2/c)}{(1-\gamma^2)\Gamma^2(1/c)}, \tag{3.62}$$

$$\delta = \int \frac{\partial \varphi^*}{\partial s^*} ss^* p_{\mathbf{s}}^0(s) = \frac{2c + (1-c)\gamma^2}{2(1-\gamma^2)}, \tag{3.63}$$

$$\rho = \int \frac{\partial \varphi^*}{\partial s} s^2 p_{\mathbf{s}}^0(s) ds = -\frac{2c - 2 + (1-3c)\gamma^2}{2(1-\gamma^2)}, \tag{3.64}$$

$$\xi = \int \frac{\partial \varphi^*}{\partial s} p_{\mathbf{s}}^0(s) ds = -\gamma\kappa, \tag{3.65}$$

$$\lambda = \int \frac{\partial^2 \varphi^*}{\partial s \partial s^*} s p_{\mathbf{s}}^0(s) ds = (c-1)\kappa, \tag{3.66}$$

$$\omega = \int \frac{\partial^2 \varphi^*}{(\partial s)^2} s p_{\mathbf{s}}^0(s) ds = -\frac{3}{2}(c-1)\gamma\kappa, \tag{3.67}$$

$$\tau = \int \frac{\partial^2 \varphi^*}{(\partial s^*)^2} s p_{\mathbf{s}}^0(s) ds = -\frac{1}{2}(c-1)\gamma\kappa. \tag{3.68}$$

Note that there exists a relationship between these parameters and the ones in the derivation of the CRB in Sect. 3.2: κ and ξ are identical. Using Corollary 2 from [35], we furthermore get

$$\delta = \eta - 1 \quad \text{and} \quad \rho = \beta - 2 \tag{3.69}$$

where $\eta = \mathrm{E}\left[|s|^2|\varphi(s)|^2\right]$ and $\beta = \mathrm{E}\left[s^2(\varphi^*(s))^2\right]$ have been defined in (3.23) and (3.25) in the previous section. These relationships hold not only for GGD but for all source distributions.
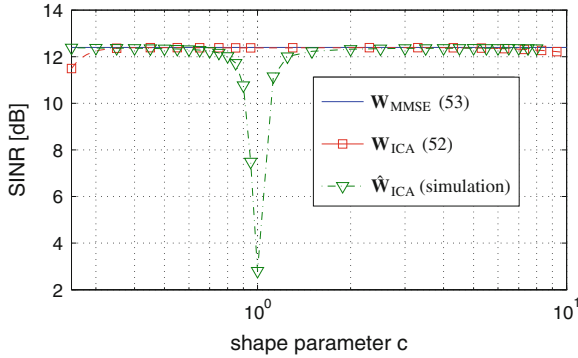
Fig. 3.5 SINR for circular complex GGD signals and circular complex noise, SNR $= 10$ dB, $L = 10^4$ samples

3.3.3.1 Circular Complex Case

For a circular complex GGD, $\gamma = 0$ and hence we get $\kappa = \frac{c^2 \Gamma(2/c)}{\Gamma^2(1/c)}, \delta = c, \rho = c-1$, $\lambda = (c-1)\kappa$ and $\xi = \omega = \tau = 0$. Figure 3.5 shows that for a wide range of the shape parameter c, both the theoretical ICA solution $\mathbf{W}_{\text{ICA}}$ and its estimate $\hat{\mathbf{W}}_{\text{ICA}}$ obtained by running KLD-ICA using $L = 10^4$ samples achieve an SINR close to that of the MMSE solution $\mathbf{W}_{\text{MMSE}}$. Note that for c close to 1, the SINR of the theoretical solution $\mathbf{W}_{\text{ICA}}$ is not achievable in practice, since all sources become Gaussian and the CRB approaches infinity for $c \to 1$ (see Sect. 3.2 and [35]). Hence estimation of $\mathbf{W}$ becomes impossible. This is reflected in Fig. 3.5: The SINR for $\hat{\mathbf{W}}_{\text{ICA}}$ estimated by KLD-ICA decreases for $c \to 1$.

Note that for strongly non-Gaussian sources ($c \ll 1$ or $c \gg 1$) the SINR of the theoretical solution $\mathbf{W}_{\text{ICA}}$ might be smaller than that for $\hat{\mathbf{W}}_{\text{ICA}}$ because $\mathbf{W}_{\text{ICA}}$ is based on a Taylor series expansion up to order σ^2. For strongly non-Gaussian sources, higher-order terms become important. These are implicitly taken into account by $\hat{\mathbf{W}}_{\text{ICA}}$ but not by $\mathbf{W}_{\text{ICA}}$.

3.3.3.2 Noncircular Complex Case

First, we study the performance with circular noise, i.e., $\mathbf{R}_{\mathbf{v}} = \mathbf{I}$ and $\bar{\mathbf{R}}_{\mathbf{v}} = \mathbf{0}$, and SNR of 10 dB. The SINR of the MMSE solution $\mathbf{W}_{\text{MMSE}}$ is 12.4 dB. Figure 3.6 shows that for a wide range of the shape parameter c and the noncircularity index γ, the theoretical ICA solution $\mathbf{W}_{\text{ICA}}$ achieves an SINR close to that of MMSE. Comparing Fig. 3.6a, b, we note that the contour plot for the simulation using $L = 10^3$ samples differs from the contour plot for the theoretical ICA solution. One reason is that for noncircular sources with the same noncircularity index $\gamma_i = \gamma$, the estimation variance increases for $c \to 1$ (see Sect. 3.2 and [35]). Hence, in the simulation the SINR decreases in the vicinity of $c = 1$. Furthermore, the smaller sample size of
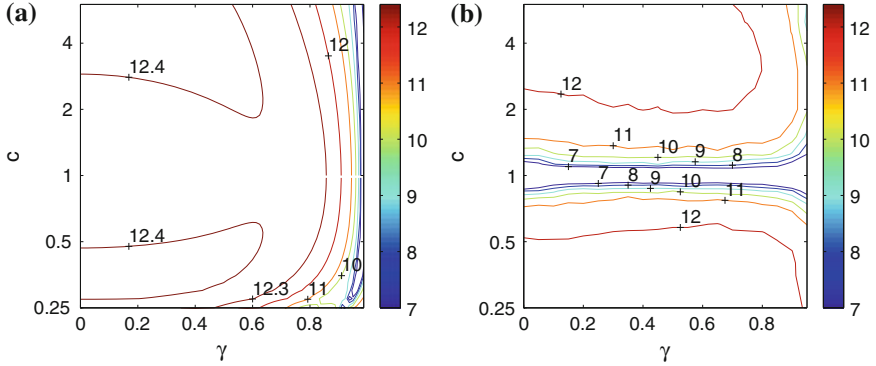
Fig. 3.6 SINR [dB] of ICA solution for noncircular complex GGD signals with $\gamma_i = \gamma$, circular complex noise and SNR = 10 dB. $\mathbf{W}_{\mathrm{ICA}}$ (52) (**a**), $\hat{\mathbf{W}}_{\mathrm{ICA}}$ (simulation, $L = 10^3$ samples) (**b**)

$L = 10^3$ leads to a larger variance of $\hat{\mathbf{W}}_{\mathrm{ICA}}$ which is not reflected in the theoretical ICA solution $\mathbf{W}_{\mathrm{ICA}}$. With a larger sample size the SINR of $\mathbf{W}_{\mathrm{ICA}}$ would be much closer to that of $\mathbf{W}_{\mathrm{MMSE}}$. However, Fig. 3.6b shows that even with a limited sample size KLD-ICA can still achieve SINR performance quite close to that of MMSE except for $c \approx 1$.

Now, we consider the case where sources are noncircular complex with $\gamma_1 = 0.5$, $\gamma_{2,3} = 0.5 \pm \Delta\gamma$ and the noise $\mathbf{v}$ is noncircular with $\mathbf{R}_{\mathbf{v}} = \mathbf{I}$ and $\bar{\mathbf{R}}_{\mathbf{v}} = 0.5 \cdot \mathbf{I}$, i.e., $\gamma_{\mathrm{noise}} = 0.5$. Figure 3.7 shows decreasing SINR values for $c \to 1$ and $\Delta\gamma \to 0$ since in that region $|\Re C_{ij}|$ in (3.100) becomes large if sources or noise are noncircular. However, except for this region, the SINR of the theoretical ICA solution (Fig. 3.7a) is still close to that of MMSE (12.4 dB). The form of the contour plot for the simulation (Fig. 3.7b) is similar to that of the theoretical solution but shows slightly lower SINR performance especially for $c \approx 1$ and small $\Delta\gamma$. This is again due to increasing estimation variance for $c \to 1$ and small $\Delta\gamma$ (see Sect. 3.2 and [35]). Nevertheless, the performance obtainable in simulations can still be considered good as long as c is not close to 1 or $\Delta\gamma$ is sufficiently large. Finally, we want to note that in Fig. 3.7 the decrease in SINR for strongly noncircular (large $\Delta\gamma$), non-Gaussian ($c \neq 1$) sources is caused by the noncircularity of the noise. The reason is that the MMSE (or maximum SINR) and the minimum KLD criterion yield different demixing matrices $\mathbf{W}$ for noncircular noise: As can be seen from (3.52), (3.97) and (3.98), $\mathbf{W}_{\mathrm{ICA}}$ depends both on the noncircularity of the sources ($\gamma_i \neq 0$) as well as on the noncircularity of the noise ($\bar{\mathbf{R}}_{\mathbf{v}} \neq \mathbf{0}$) whereas $\mathbf{W}_{\mathrm{MMSE}}$ from (3.54) only depends on the normal covariance matrix of the noise $\mathbf{R}_{\mathbf{v}}$. This is due to the different cost functions: Minimization of KLD makes the pdf of the demixed signals as similar to the assumed pdf of the sources as possible whereas MMSE minimizes the expected quadratic error between the demixed signals and the original sources. For circular noise, the difference between $\mathbf{W}_{\mathrm{ICA}}$ and $\mathbf{W}_{\mathrm{MMSE}}$ in terms of SINR is much smaller.
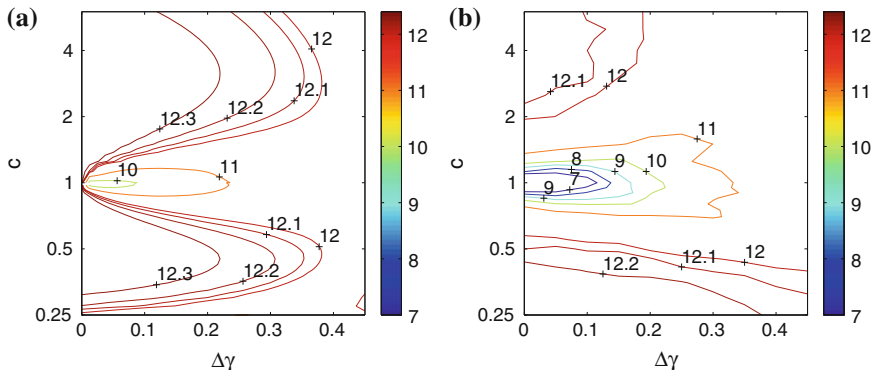
Fig. 3.7 SINR [dB] of ICA solution for noncircular complex GGD signals with $\gamma_1 = 0.5$, $\gamma_{2,3} = \gamma_1 \pm \Delta\gamma$, noncircular complex noise, and SNR $= 10\,$dB. $\mathbf{W}_{\mathrm{ICA}}$ (52) (**a**), $\hat{\mathbf{W}}_{\mathrm{ICA}}$ (simulation, $L = 10^3$ samples) (**b**)
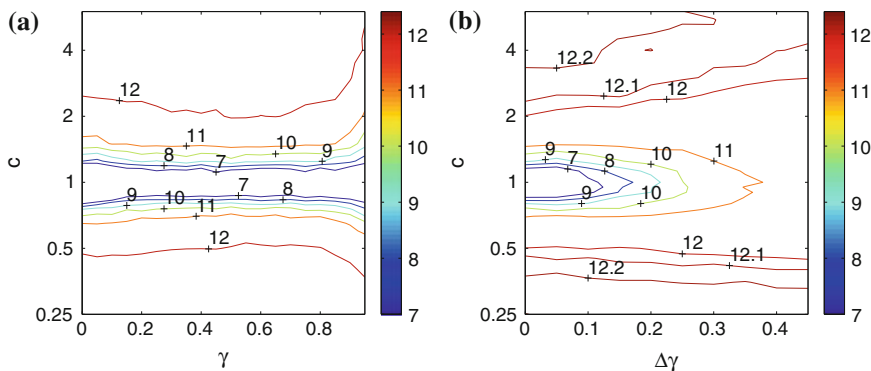

Fig. 3.8 SINR [dB] of ICA-EBM solution for noncircular complex GGD signals, $L = 10^3$ samples and SNR $= 10\,$dB. $\gamma_i = \gamma$, circular complex noise(**a**), $\gamma_1 = 0.5$, $\gamma_{2,3} = \gamma_1 \pm \delta\gamma$ (**b**)

In summary, the results in this subsection have shown that

- in many cases the theoretical solution $\mathbf{W}_{\mathrm{ICA}}$ of KLD-ICA can achieve an SINR close to the optimum attainable by the MMSE demixing matrix $\mathbf{W}_{\mathrm{MMSE}}$.
- for sources following a GGD, $\hat{\mathbf{W}}_{\mathrm{ICA}}$ obtained by running KLD-ICA with a finite amount of samples L can achieve an SINR quite close to that of $\mathbf{W}_{\mathrm{MMSE}}$ except for (nearly) Gaussian sources with similar noncircularity indices.
- for strongly noncircular, non-Gaussian sources and noncircular noise, the minimization of the KLD and of the MSE yield different solutions.

Although we assumed that we perfectly know the distributions of the sources, other approaches such as ICA-EBM [30] exist which do not require such knowledge. As shown in Fig. 3.8, simulation results using ICA-EBM show similar SINR performance as KLD-ICA (see Figs. 3.6b, 3.7b).

3.3.4 Extension to Overdetermined Case

ICA algorithms for the overdetermined case have already been studied in a number
of publications (see e.g., [11, 25, 49, 50]). In the overdetermined case, $\mathbf{x} \in \mathbb{C}^M$ with
$M > N$. In the noiseless case we can select any N rows of $\mathbf{x}$ to perform ICA as
long as the corresponding square mixing matrix $\widetilde{\mathbf{A}}$ is invertible. When we consider
noisy mixtures, this does not hold since the information contained in the $M - N$
additional observations is useful to improve demixing. Hence, we need to consider
the KLD for $M > N$. In this case, the demixing matrix $\mathbf{W}$ can be decomposed as
$\mathbf{W} = \begin{bmatrix} \mathbf{W}_1 & \mathbf{W}_2 \end{bmatrix}$, where $\mathbf{W}_1 \in \mathbb{C}^{N \times N}$ and $\mathbf{W}_2 \in \mathbb{C}^{N \times (M-N)}$. We define an auxiliary
vector $\bar{\mathbf{y}} \in \mathbb{C}^M$:

$$\bar{\mathbf{y}} = \begin{bmatrix} \mathbf{W}_1 & \mathbf{W}_2 \\ \mathbf{0} & \mathbf{I}_{M-N} \end{bmatrix} \begin{bmatrix} \mathbf{x}_1 \\ \mathbf{x}_2 \end{bmatrix} = \check{\mathbf{W}}\mathbf{x} = \begin{bmatrix} \mathbf{y} \\ \mathbf{x}_2 \end{bmatrix} \tag{3.70}$$

Then we calculate $p_{\mathbf{y}}(\mathbf{y}; \mathbf{W})$ by

$$p_{\bar{\mathbf{y}}}(\bar{\mathbf{y}}; \check{\mathbf{W}}) = \frac{1}{|\det(\check{\mathbf{W}})|^2} p_{\mathbf{x}}(\mathbf{x}) = \frac{1}{|\det(\mathbf{W}_1)|^2} p_{\mathbf{x}}(\mathbf{x}), \tag{3.71}$$

$$p_{\mathbf{y}}(\mathbf{y}; \mathbf{W}) = \frac{1}{|\det(\mathbf{W}_1)|^2} \int p_{\mathbf{x}}(\mathbf{x}) d\mathbf{x}_2. \tag{3.72}$$

since the linear transformation of a complex random vector yields $|\det(\mathbf{W})|^2$ instead
of $|\det(\mathbf{W})|$ in the real case (see [4, 42]).

Using the above steps we obtain the modified KLD for $M > N$

$$D_{\mathrm{KL}}(\mathbf{W}) = -\ln|\det(\mathbf{W}_1)|^2 - \sum_{i=1}^{N} \mathrm{E}\left[\ln p_{s_i}(y_i)\right] + \text{const.} \tag{3.73}$$

instead of $D_{\mathrm{KL}}(\mathbf{W}) = -\ln|\det(\mathbf{W})|^2 - \sum_{i=1}^{N} \mathrm{E}\left[\ln p_{s_i}(y_i)\right] + \text{const. for } M = N$.

To derive $\mathbf{W}_{\mathrm{ICA}}$ for $M > N$, we could now perform a similar Taylor series expan-
sion as for $M = N$. However, it is more convenient to reduce the overdetermined
case $M > N$ to the determined case by applying a linear transform to the data to
condense all information about the source signals in the first N observations and by
applying another transform to decorrelate the noise terms in the first N observations
from those in the remaining $M - N$ observations. The result of these two transforms
has a similar effect as a dimension reduction using principal component analysis
(PCA) except that the correlation matrix of the observations is only block-diagonal
instead of diagonal. To derive $\mathbf{W}_{\mathrm{ICA}}$, we can then combine the solution for the deter-
mined case with the linear transforms. Note that this approach is only used for the
analysis of KLD-based ICA for the overdetermined case because it simplifies the
theoretical derivation. In ICA applications, the transforms are done implicitly by the
algorithm itself.

The first step of this procedure is to use the orthogonal transform $\mathbf{Q}$ defined by the decomposition $\mathbf{A} = \mathbf{Q}^H \begin{bmatrix} \bar{\mathbf{A}}_1 \\ \mathbf{0} \end{bmatrix}$ to condense all information about the source signals in the first N observations:

$$
\begin{aligned}
\bar{\mathbf{x}} = \mathbf{Q}\mathbf{x} &= \mathbf{Q}(\mathbf{A}\mathbf{s} + \mathbf{v}) \\
&= \mathbf{Q}\mathbf{Q}^H \begin{bmatrix} \bar{\mathbf{A}}_1 \\ \mathbf{0} \end{bmatrix} \mathbf{s} + \mathbf{Q}\mathbf{v} = \begin{bmatrix} \bar{\mathbf{A}}_1 \\ \mathbf{0} \end{bmatrix} \mathbf{s} + \bar{\mathbf{v}} = \begin{bmatrix} \bar{\mathbf{A}}_1\mathbf{s} \\ \mathbf{0} \end{bmatrix} + \begin{bmatrix} \bar{\mathbf{v}}_1 \\ \bar{\mathbf{v}}_2 \end{bmatrix} = \begin{bmatrix} \bar{\mathbf{x}}_1 \\ \bar{\mathbf{x}}_2 \end{bmatrix},
\end{aligned}
\tag{3.74}
$$

$$
\mathbf{R}_{\bar{\mathbf{v}}} = \frac{1}{\sigma^2}\mathrm{E}\left(\bar{\mathbf{v}}\bar{\mathbf{v}}^H\right) = \mathbf{Q}\mathbf{R}_{\mathbf{v}}\mathbf{Q}^H = \begin{bmatrix} \mathbf{R}_{\bar{\mathbf{v}}_{11}} & \mathbf{R}_{\bar{\mathbf{v}}_{12}} \\ \mathbf{R}_{\bar{\mathbf{v}}_{21}} & \mathbf{R}_{\bar{\mathbf{v}}_{22}} \end{bmatrix}.
\tag{3.75}
$$

Note that $\bar{\mathbf{v}}_1$ and $\bar{\mathbf{v}}_2$ may be correlated, i.e., $\bar{\mathbf{x}}_2 = \bar{\mathbf{v}}_2$ is useful for the processing of $\bar{\mathbf{x}}_1 = \bar{\mathbf{A}}_1\mathbf{s} + \bar{\mathbf{v}}_1$ to reduce the impact of $\bar{\mathbf{v}}_1$. Hence, we decorrelate the noise terms $\bar{\mathbf{v}}_1$ and $\bar{\mathbf{v}}_2$ by a second transform $\mathbf{T} = \begin{bmatrix} \mathbf{I}_N & -\mathbf{R}_{\bar{\mathbf{v}}_{12}}\mathbf{R}_{\bar{\mathbf{v}}_{22}}^{-1} \\ \mathbf{0} & \mathbf{I}_{M-N} \end{bmatrix}$:

$$
\tilde{\mathbf{x}} = \mathbf{T}\bar{\mathbf{x}} = \mathbf{T}\begin{bmatrix} \bar{\mathbf{A}}_1\mathbf{s} \\ \mathbf{0} \end{bmatrix} + \mathbf{T}\bar{\mathbf{v}} = \begin{bmatrix} \bar{\mathbf{A}}_1\mathbf{s} \\ \mathbf{0} \end{bmatrix} + \tilde{\mathbf{v}} = \begin{bmatrix} \bar{\mathbf{A}}_1\mathbf{s} \\ \mathbf{0} \end{bmatrix} + \begin{bmatrix} \tilde{\mathbf{v}}_1 \\ \tilde{\mathbf{v}}_2 \end{bmatrix} = \begin{bmatrix} \tilde{\mathbf{x}}_1 \\ \tilde{\mathbf{x}}_2 \end{bmatrix},
\tag{3.76}
$$

$$
\mathbf{R}_{\tilde{\mathbf{v}}} = \frac{1}{\sigma^2}\mathrm{E}\left(\tilde{\mathbf{v}}\tilde{\mathbf{v}}^H\right) = \begin{bmatrix} \mathbf{R}_{\bar{\mathbf{v}}_{11}} - \mathbf{R}_{\bar{\mathbf{v}}_{12}}\mathbf{R}_{\bar{\mathbf{v}}_{22}}^{-1}\mathbf{R}_{\bar{\mathbf{v}}_{21}} & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_{\bar{\mathbf{v}}_{22}} \end{bmatrix}.
\tag{3.77}
$$

$\tilde{\mathbf{v}}_1$ and $\tilde{\mathbf{v}}_2$ are now uncorrelated and $\tilde{\mathbf{x}}_2 = \tilde{\mathbf{v}}_2$ does not contain any second-order information useful for the processing of $\tilde{\mathbf{x}}_1 = \bar{\mathbf{A}}_1\mathbf{s} + \tilde{\mathbf{v}}_1$.

The separated signals $\mathbf{y}$ are now obtained by

$$
\mathbf{y} = \mathbf{W}\mathbf{x} = \mathbf{W}\mathbf{Q}^H\mathbf{T}^{-1}\tilde{\mathbf{x}} = \tilde{\mathbf{W}}\tilde{\mathbf{x}} = \begin{bmatrix} \tilde{\mathbf{W}}_1 & \tilde{\mathbf{W}}_2 \end{bmatrix}\begin{bmatrix} \tilde{\mathbf{x}}_1 \\ \tilde{\mathbf{x}}_2 \end{bmatrix} = \tilde{\mathbf{W}}_1\tilde{\mathbf{x}}_1 + \tilde{\mathbf{W}}_2\tilde{\mathbf{x}}_2 = \mathbf{y}_1 + \mathbf{y}_2
\tag{3.78}
$$

with $\tilde{\mathbf{W}} = \mathbf{W}\mathbf{Q}^H\mathbf{T}^{-1}$. The noise-only contribution $\mathbf{y}_2 = \tilde{\mathbf{W}}_2\tilde{\mathbf{x}}_2 = \tilde{\mathbf{W}}_2\tilde{\mathbf{v}}_2$ to $\mathbf{y}$ is uncorrelated to $\mathbf{y}_1 = \tilde{\mathbf{W}}_1\tilde{\mathbf{x}}_1$. Hence, it is sufficient to consider the first N observations $\tilde{\mathbf{x}}_1$ to derive the ICA solution for $\tilde{\mathbf{W}}_1$.

Considering the KLD (3.73) for the transformed demixing model $\mathbf{y} = \begin{bmatrix} \tilde{\mathbf{W}}_1 & \tilde{\mathbf{W}}_2 \end{bmatrix} \begin{bmatrix} \tilde{\mathbf{x}}_1 \\ \tilde{\mathbf{x}}_2 \end{bmatrix}$, we get

$$
D_{\mathrm{KL}}(\tilde{\mathbf{W}}) = -\ln|\det(\tilde{\mathbf{W}}_1)| - \sum_{i=1}^{N}\mathrm{E}\left[\ln p_{s_i}(y_i)\right] + \mathrm{const.}
\tag{3.79}
$$

with $\tilde{\mathbf{W}} = \begin{bmatrix} \tilde{\mathbf{W}}_1 & \tilde{\mathbf{W}}_2 \end{bmatrix}$. The real derivatives of $D_{\mathrm{KL}}(\tilde{\mathbf{W}})$ with respect to $\tilde{\mathbf{W}}_1$ and $\tilde{\mathbf{W}}_2$ are given by

$$\frac{\partial D_{\mathrm{KL}}(\widetilde{\mathbf{W}})}{\partial \widetilde{\mathbf{W}}_1} = \mathrm{E}\left[\boldsymbol{\varphi}(\mathbf{y})\widetilde{\mathbf{x}}_1^H\right] - \widetilde{\mathbf{W}}_1^{-H} = \mathrm{E}\left[\boldsymbol{\varphi}(\mathbf{y})\widetilde{\mathbf{y}}_1^H - \mathbf{I}\right]\widetilde{\mathbf{W}}_1^{-H} \overset{!}{=} \mathbf{0} \qquad (3.80)$$

$$\frac{\partial D_{\mathrm{KL}}(\widetilde{\mathbf{W}})}{\partial \widetilde{\mathbf{W}}_2} = \mathrm{E}\left[\boldsymbol{\varphi}(\mathbf{y})\widetilde{\mathbf{x}}_2^H\right] = \mathrm{E}\left[\boldsymbol{\varphi}(\mathbf{y})\widetilde{\mathbf{y}}_2^H\right]\widetilde{\mathbf{W}}_2^{-H} \overset{!}{=} \mathbf{0} \qquad (3.81)$$

A perturbation analysis of (3.81) at $\mathbf{y} = \widetilde{\mathbf{W}}_1\bar{\mathbf{A}}_1\mathbf{s}$ yields $\widetilde{\mathbf{W}}_2 = \mathcal{O}(\sigma^4)$ due to $\widetilde{\mathbf{x}}_2 = \widetilde{\mathbf{v}}_2$. Hence, $\mathbf{y}$ is given by $\mathbf{y} = \widetilde{\mathbf{W}}_1\widetilde{\mathbf{x}}_1 + \mathcal{O}(\sigma^4)$. The solution for $\widetilde{\mathbf{W}}_1$ is similar to the case of $M = N$

$$\widetilde{\mathbf{W}}_{1,\mathrm{ICA}} = (\mathbf{I} + \sigma^2\mathbf{C})\bar{\mathbf{A}}_1^{-1} + \mathcal{O}(\sigma^4) \qquad (3.82)$$

where the elements of $\mathbf{C}$ can be computed from (3.97) and (3.98) with $\bar{\mathbf{R}}_{-1} = \mathbf{0}$ and $\mathbf{R}_{-1} = (\bar{\mathbf{A}}_1^{-1}(\mathbf{R}_{\bar{\mathbf{v}}_{11}} - \mathbf{R}_{\bar{\mathbf{v}}_{12}}\mathbf{R}_{\bar{\mathbf{v}}_{22}}^{-1})\bar{\mathbf{A}}_1^{-H})$.

Finally, we need to combine $\widetilde{\mathbf{W}}_{1,\mathrm{ICA}}$, $\mathbf{T}$ and $\mathbf{Q}$ to form the final solution:

$$\widetilde{\mathbf{W}}_{\mathrm{ICA}} = \left[\widetilde{\mathbf{W}}_{1,\mathrm{ICA}} \ \mathbf{0}\right] + \mathcal{O}(\sigma^4),$$

$$\mathbf{W}_{\mathrm{ICA}} = \widetilde{\mathbf{W}}_{\mathrm{ICA}}\mathbf{T}\mathbf{Q} = \widetilde{\mathbf{W}}_{1,\mathrm{ICA}}\left[\mathbf{I} - \mathbf{R}_{\bar{\mathbf{v}}_{12}}\mathbf{R}_{\bar{\mathbf{v}}_{22}}^{-1}\right]\mathbf{Q} + \mathcal{O}(\sigma^4). \qquad (3.83)$$

Note that for the case of noncircular complex noise, the presented transformation does not work since we would need to take into account the pseudo-covariance matrix of the noise.

3.3.4.1 Results for Circular Complex GGD

Here, we study the performance for the overdetermined case with $M = 6$ sensors, $N = 3$ sources, and circular complex noise. The sources follow a circular complex GGD distribution with identical shape parameters c. Similar to Sect. 3.3.3, we use the mixing matrix $\mathbf{A} = [a_{mn}]$ with $a_{mn} = e^{-j\pi m \sin\theta_n}$ with $\theta_n = -60°, 0°, 60°$. We first consider spatially uncorrelated noise with $\mathbf{R_v} = \mathbf{I}$. Figure 3.9a shows that for a wide range of the shape parameter c, both the theoretical ICA solution $\mathbf{W}_{\mathrm{ICA}}$ and its estimate $\hat{\mathbf{W}}_{\mathrm{ICA}}$ obtained by running KLD-ICA using $L = 10^4$ samples achieve an SINR close to that of the MMSE solution $\mathbf{W}_{\mathrm{MMSE}}$. Furthermore, note that additional sensors can improve the SINR of the demixed signals: Using only the first $M = 3$ sensors, $\mathbf{W}_{\mathrm{MMSE}}$ achieves an SINR of 12.4 dB (see Fig. 3.5), whereas with $M = 6$ sensors it achieves an SINR of 17.4 dB.

When the noise $\mathbf{v}$ is correlated with the normalized correlation matrix

$$\mathbf{R_v} = \begin{bmatrix} 1.00+0.00j & 0.62+0.23j & 0.44-0.16j & 0.46+0.11j & -0.09+0.26j & -0.03+0.09j \\ 0.62-0.23j & 1.00+0.00j & 0.56+0.06j & 0.47-0.13j & 0.44+0.18j & -0.09+0.26j \\ 0.44+0.16j & 0.56-0.06j & 1.00+0.00j & 0.52+0.09j & 0.47-0.13j & 0.46+0.11j \\ 0.46-0.11j & 0.47+0.13j & 0.52-0.09j & 1.00+0.00j & 0.56+0.06j & 0.44-0.16j \\ -0.09-0.26j & 0.43-0.18j & 0.47+0.13j & 0.56-0.06j & 1.00+0.00j & 0.62+0.23j \\ -0.03-0.09j & -0.09-0.26j & 0.46-0.11j & 0.44+0.16j & 0.62-0.23j & 1.00+0.00j \end{bmatrix},$$
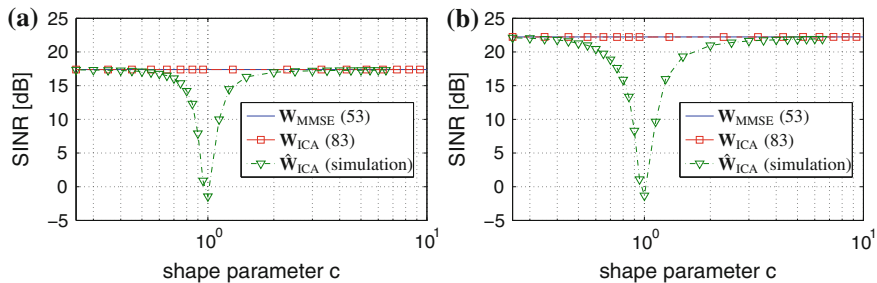
Fig. 3.9 SINR for overdetermined case with circular complex GGD signals and circular complex noise, SNR = 10 dB, $L = 10^4$ samples. $\mathbf{R_v} = \mathbf{I}$ **(a)**, $\mathbf{R_v} \neq \mathbf{I}$ **(b)**

$\mathbf{W}_{\text{MMSE}}$ achieves an SINR of 22.2 dB for an SNR of 10 dB and all $M = 6$ sensors. With the first $M = 3$ sensors, it achieves only an SINR of 13.3 dB. Compared to the case of uncorrelated noise, the form of the SINR curve for $\hat{\mathbf{W}}_{\text{ICA}}$ changes slightly but it is still quite close to that of $\mathbf{W}_{\text{MMSE}}$ except for $c \approx 1$ (see Fig. 3.9b).

3.4 Conclusion

We have derived an analytic expression for the demixing matrix of KLD-based ICA for the low noise regime. We have considered the general noncircular complex determined case. The solution for the circular complex and real case can be derived as special cases. Furthermore, we have shown how to reduce the overdetermined case $M > N$ to the determined case. Although the KLD and MMSE solutions differ, linear demixing based on these two criteria yields demixed signals with similar SINR in many cases. In practice, however, not only the bias studied in this chapter but also the variance of the estimate are important for SINR. For the noiseless case, the variance of the estimated demixing matrix is lower bounded by the CRB derived in Sect. 3.2 and [35].

Appendix 1

Values of κ, ξ, β, η for Complex GGD

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

$$p(s, s^*) = \frac{c\alpha \cdot \exp\left(-\left[\frac{\alpha/2}{\gamma^2-1}\left(\gamma s^2 + \gamma s^{*2} - 2ss^*\right)\right]^c\right)}{\pi\, \Gamma(1/c)(1 - \gamma^2)^{1/2}}, \qquad (3.84)$$

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

$$\varphi(s, s^*) = \frac{2c(\alpha/2)^c}{(\gamma^2 - 1)^c} \left(\gamma s^2 + \gamma (s^*)^2 - 2ss^* \right)^{c-1} (\gamma s^* - s). \tag{3.85}$$

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

$$\kappa = \mathrm{E}\left[|\varphi(s)|^2 \right] = \frac{c^2 \Gamma(2/c)}{(1 - \gamma^2)\Gamma^2(1/c)}, \tag{3.86}$$

$$\xi = \mathrm{E}\left[(\varphi^*(s))^2 \right] = -\frac{c^2 \gamma \Gamma(2/c)}{(1 - \gamma^2)\Gamma^2(1/c)} = -\gamma\kappa, \tag{3.87}$$

$$\eta = \mathrm{E}\left[|s|^2 |\varphi(s)|^2 \right] = \frac{(c + 1) \cdot (2 - \gamma^2)}{2(1 - \gamma^2)}, \tag{3.88}$$

$$\beta = \mathrm{E}\left[s^2 (\varphi^*(s))^2 \right] = \frac{(c + 1) \cdot (2 - 3\gamma^2)}{2(1 - \gamma^2)}. \tag{3.89}$$

Induced CRB for Real ICA

Here, we briefly review the iCRB for real ICA [41, 45]. In the following, all real quantities q are denoted as $\mathring{q}$. In the derivation of the iCRB for the real case $\mathring{\varphi}(\mathring{s}) = -\partial \ln p(\mathring{s})/\partial \mathring{s}$ and the parameters $\mathring{\kappa} = E[\mathring{\varphi}^2(\mathring{s})]$, $\mathring{\eta} = \mathrm{E}[\mathring{s}^2 \mathring{\varphi}^2(\mathring{s})] = 2 + \mathrm{E}\left[\mathring{s}^2 \frac{\partial \mathring{\varphi}(\mathring{s})}{\partial \mathring{s}} \right]$ are defined using real derivatives. In [41, 45] it was shown that

$$\mathrm{var}(\hat{G}_{ii}) \geq \frac{1}{L(\mathring{\eta}_i - 1)}, \tag{3.90}$$

$$\mathrm{var}(\hat{G}_{ij}) \geq \frac{1}{L} \frac{\mathring{\kappa}_j}{\mathring{\kappa}_i \mathring{\kappa}_j - 1}. \tag{3.91}$$

Appendix 2

Here we derive an analytic expression for $\mathbf{W}_{\mathrm{ICA}}$ in the presence of noise by using a perturbation analysis. Motivated by $\mathbf{W}_{\mathrm{ICA}} \overset{\sigma^2 = 0}{=} \mathbf{A}^{-1}$, we assume that $\mathbf{W}_{\mathrm{ICA}}$ can be written as $\mathbf{W}_{\mathrm{ICA}} = \mathbf{A}^{-1} + \sigma^2 \mathbf{B} + \mathcal{O}(\sigma^4)$ and derive $\mathbf{B}$ by a two-step perturbation analysis:

1. Taylor series approximation of $\mathrm{E}(\boldsymbol{\varphi}^*(\mathbf{y})\mathbf{y}^T)$ in (3.51) at $\mathbf{y} = \hat{\mathbf{y}} = \mathbf{W}_{\mathrm{ICA}}\mathbf{As}$,
2. Taylor series approximation of the result of the above step by exploiting $\mathbf{W}_{\mathrm{ICA}} = \mathbf{A}^{-1} + \sigma^2 \mathbf{B} + \mathcal{O}(\sigma^4)$ and $\hat{\mathbf{y}} = \mathbf{s} + \sigma^2 \mathbf{BAs} + \mathcal{O}(\sigma^4) = \mathbf{s} + \sigma^2 \mathbf{Cs} + \mathcal{O}(\sigma^4) = \mathbf{s} + \sigma^2 \mathbf{b} + \mathcal{O}(\sigma^4)$ with $\mathbf{C} = \mathbf{BA}$ and $\mathbf{b} = \mathbf{Cs} = [b_1, \ldots, b_N]^T$.

In this way, we determine explicitly the deviation $\sigma^2 \mathbf{B}$ of $\mathbf{W}_{\text{ICA}}$ from the inverse solution $\mathbf{A}^{-1}$.

The general Taylor series expansion of $\varphi^*(y) \widehat{=} \varphi^*(y, y^*)$ is given as

$$\varphi^*(y, y^*) = \varphi^*(\hat{y}, \hat{y}^*) + \frac{\partial \varphi^*}{\partial y} \Delta y + \frac{\partial \varphi^*}{\partial y^*} \Delta y^* + \frac{1}{2} \left(\frac{\partial^2 \varphi^*}{(\partial y)^2} (\Delta y)^2 + \frac{\partial^2 \varphi^*}{(\partial y^*)^2} (\Delta y^*)^2 \right)$$

$$+ \frac{\partial^2 \varphi^*}{\partial y \partial y^*} \Delta y \Delta y^* + \dots$$

$$= \varphi^*(\hat{y}, \hat{y}^*) + \varpi(y, y^*) \Delta y + \vartheta(y, y^*) \Delta y^*$$

$$+ \frac{1}{2} \left(\nu(y, y^*)(\Delta y)^2 + \zeta(y, y^*)(\Delta y^*)^2 \right) + \epsilon(y, y^*) \Delta y \Delta y^* + \dots \quad (3.92)$$

with $\varpi(y, y^*) = \frac{\partial \varphi^*}{\partial y}$, $\vartheta(y, y^*) = \frac{\partial \varphi^*}{\partial y^*}$, $\nu(y, y^*) = \frac{\partial^2 \varphi^*}{(\partial y)^2}$, $\zeta(y, y^*) = \frac{\partial^2 \varphi^*}{(\partial y^*)^2}$ and $\epsilon(y, y^*) = \frac{\partial^2 \varphi^*}{\partial y \partial y^*}$. To simplify notation, we will drop the dependence of $\varphi^*(\cdot)$, $\varpi(\cdot)$, $\vartheta(\cdot)$, $\nu(\cdot)$, $\zeta(\cdot)$, $\epsilon(\cdot)$ on y^* and keep only the dependence on y in the following.

Let

$$\rho_i = \mathrm{E}\left[\varpi_i(s_i) s_i^2\right], \qquad \delta_i = \mathrm{E}\left[\vartheta_i(s_i) s_i^* s_i\right], \tag{3.93}$$

$$\kappa_i = \mathrm{E}\left[\vartheta_i(s_i)\right], \qquad \xi_i = \mathrm{E}\left[\varpi_i(s_i)\right], \tag{3.94}$$

$$\omega_i = \mathrm{E}\left[\nu_i(s_i) s_i\right], \qquad \tau_i = \mathrm{E}\left[\zeta_i(s_i) s_i\right], \tag{3.95}$$

$$\lambda_i = \mathrm{E}\left[\epsilon_i(s_i) s_i\right], \qquad \gamma_i = \mathrm{E}[s_i^2]. \tag{3.96}$$

As shown in [31, 34], $\mathbf{W}_{\text{ICA}} = \mathbf{A}^{-1} + \sigma^2 \mathbf{C}$, where the elements of $\mathbf{C}$ can be computed from

$$\rho_i C_{ii} + \delta_i C_{ii}^* + C_{ii} = -(\kappa_i + \lambda_i) \left[\mathbf{R}_{-1}\right]_{ii} - (\xi_i + \frac{1}{2}\omega_i) \left[\bar{\mathbf{R}}_{-1}\right]_{ii} - \frac{1}{2}\tau_i \left[\bar{\mathbf{R}}_{-1}\right]_{ii}^*. \tag{3.97}$$

and

$$\gamma_j \xi_i C_{ij} + \kappa_i C_{ij}^* + C_{ji} = -\kappa_i \left[\mathbf{R}_{-1}\right]_{ij}^* - \xi_i \left[\bar{\mathbf{R}}_{-1}\right]_{ij},$$

$$\gamma_i \xi_j C_{ji} + \kappa_j C_{ji}^* + C_{ij} = -\kappa_j \left[\mathbf{R}_{-1}\right]_{ji}^* - \xi_j \left[\bar{\mathbf{R}}_{-1}\right]_{ji}. \tag{3.98}$$

with the transformed noise covariance matrix $\mathbf{R}_{-1} = \mathbf{W}\mathbf{R}_{\mathbf{v}}\mathbf{W}^H = \mathbf{A}^{-1}\mathbf{R}_{\mathbf{v}}\mathbf{A}^{-H} + \mathcal{O}(\sigma^2)$ and the transformed noise pseudo-covariance matrix $\bar{\mathbf{R}}_{-1} = \mathbf{W}\bar{\mathbf{R}}_{\mathbf{v}}\mathbf{W}^T = \mathbf{A}^{-1}\bar{\mathbf{R}}_{\mathbf{v}}\mathbf{A}^{-T} + \mathcal{O}(\sigma^2)$. Note that $\mathbf{R}_{-1}^H = \mathbf{R}_{-1}$ and $\bar{\mathbf{R}}_{-1}^T = \bar{\mathbf{R}}_{-1}$.

If $p(s, s^*)$ is symmetric in the real part $\Re s$ or imaginary part $\Im s$ of s, i.e., $p(-\Re s, \Im s) = p(\Re s, \Im s)$ or $p(\Re s, -\Im s) = p(\Re s, \Im s)$, the parameters $\kappa_i, \rho_i, \delta_i, \lambda_i, \xi_i, \omega_i, \tau_i$ are real. For $\rho_i + 1 \pm \delta_i \neq 0$, we then get from (3.97)

$$\Re C_{ii} = -\frac{(\kappa_i + \lambda_i)\left[\mathbf{R}_{-1}\right]_{ii} + (\xi_i + \frac{1}{2}(\omega_i + \tau_i))\left[\Re\bar{\mathbf{R}}_{-1}\right]_{ii}}{\rho_i + 1 + \delta_i},$$

$$\Im C_{ii} = -\frac{(\xi_i + \frac{1}{2}(\omega_i - \tau_i))\left[\Im\bar{\mathbf{R}}_{-1}\right]_{ii}}{\rho_i + 1 - \delta_i}. \qquad (3.99)$$

For $(\gamma_j\xi_i + \kappa_i)(\gamma_i\xi_j + \kappa_j) \neq 1$ and $(\gamma_j\xi_i - \kappa_i)(\gamma_i\xi_j - \kappa_j) \neq 1$, we obtain from (3.98)

$$\Re C_{ij} = \frac{(\kappa_j - \kappa_i(\gamma_i\xi_j + \kappa_j))\left[\Re\mathbf{R}_{-1}\right]_{ij} + (\xi_j - \xi_i(\gamma_i\xi_j + \kappa_j))\left[\Re\bar{\mathbf{R}}_{-1}\right]_{ij}}{(\gamma_j\xi_i + \kappa_i)(\gamma_i\xi_j + \kappa_j) - 1},$$

$$\Im C_{ij} = \frac{(\kappa_j + \kappa_i(\gamma_i\xi_j - \kappa_j))\left[\Im\mathbf{R}_{-1}\right]_{ij} + (\xi_j - \xi_i(\gamma_i\xi_j - \kappa_j))\left[\Im\bar{\mathbf{R}}_{-1}\right]_{ij}}{(\gamma_j\xi_i - \kappa_i)(\gamma_i\xi_j - \kappa_j) - 1}.$$

$$(3.100)$$

References

1. Adali, T., Li, H.: A practical formulation for computation of complex gradients and its application to maximum likelihood ICA. In: Proc. IEEE Conference on Acoustics, Speech, and Signal Processing (ICASSP), vol. 2, pp. II-633-II-636 (2007)
2. Adali, T., Li, H.: Complex-valued adaptive signal processing, ch. 1. In: T. Adali, S. Haykin (eds.) Adaptive Signal Processing: Next Generation Solutions, pp. 1–85. Wiley, New York (2010)
3. Adali, T., Li, H., Novey, M., Cardoso, J.F.: Complex ica using nonlinear functions. IEEE Trans. Signal Process. **56**(9), 4536–4544 (2008)
4. Adali, T., Schreier, P., Scharf, L.: Complex-valued signal processing: the proper way to deal with impropriety. IEEE Trans. Signal Process. **59**(11), 5101–5125 (2011)
5. Anderson, M., Li, X.L., Rodriquez, P.A., Adali, T.: An effective decoupling method for matrix optimization and its application to the ICA problem. In: Proceedings of the IEEE Conference on Acoustics, Speech, and Signal Processing (ICASSP), pp. 1885–1888 (2012)
6. Brandwood, D.H.: A complex gradient operator and its application in adaptive array theory. IEE Proc. **130**, 11–16 (1983)
7. Cardoso, J., Souloumiac, A.: Blind beamforming for non-gaussian signals. Radar Signal Process. IEE Proc. F **140**(6), 362–370 (1993)
8. Cardoso, J.F.: On the performance of orthogonal source separation algorithms. In: Proceedings of the European Signal Processing Conference (EUSIPCO), pp. 776–779 (1994)
9. Cardoso, J.F.: Blind signal separation: statistical principles. Proc. IEEE **86**(10), 2009–2025 (1998)
10. Cardoso, J.F., Adali, T.: The maximum likelihood approach to complex ICA. In: Proceedings of the IEEE Conference on Acoustics, Speech, and Signal Processing (ICASSP), vol. 5, pp. 673–676 (2006)
11. Cichocki, A., Sabala, I., Choi, S., Orsier, B., Szupiluk, R.: Self adaptive independent component analysis for sub-gaussian and super-gaussian mixtures with unknown number of sources and additive noise. In: Proceedings of 1997 International Symposium on Nonlinear Theory and its Applications (NOLTA-97), vol. 2, pp. 731–734 (1997)
12. Comon, P., Jutten, C. (eds.): Handbook of Blind Source Separation: Independent Component Analysis and Applications, 1st edn. Elsevier, Amsterdam (2010)
13. Davies, M.: Identifiability issues in noisy ica. IEEE Signal Process. Lett. **11**(5), 470–473 (2004)

14. De Lathauwer, L., De Moor, B.: On the blind separation of non-circular sources. In: Proceedings of the European Signal Processing Conference (EUSIPCO), vol. 2, pp. 99–102. Toulouse, France (2002)

15. Doron, E., Yeredor, A., Tichavsky, P.: Cramér-Rao-induced bound for blind separation of stationary parametric gaussian sources. IEEE Signal Process. Lett. **14**(6), 417–420 (2007)

16. Douglas, S., Cichocki, A., Amari, S.: A bias removal technique for blind source separation with noisy measurements. Eletron. Lett. **34**(14), 1379–1380 (1998)

17. Douglas, S.C.: Fixed-point algorithms for the blind separation of arbitrary complex-valued non-gaussian signal mixtures. EURASIP J. Appl. Signal Process. **2007**(1), Article ID 36,525 (2007)

18. Eriksson, J., Koivunen, V.: Complex-valued ICA using second order statistics. pp. 183–192 (2004)

19. Eriksson, J., Koivunen, V.: Complex random vectors and ICA models: identifiability, uniqueness, and separability. IEEE Trans. Inf. Theory **52**(3), 1017–1029 (2006)

20. Fiori, S.: Neural independent component analysis by maximum-mismatch learning principle. Neural Netw. **16**(8), 1201–1221 (2003)

21. Hjørungnes, A.: Complex-Valued Matrix Derivatives. Cambridge University Press, Cambridge (2011)

22. Horn, R.A., Johnson, C.R.: Matrix analysis, 1st publis. (1985), 10th print edn. Cambridge University Press, Cambridge (1999)

23. Hyvärinen, A.: Independent component analysis in the presence of gaussian noise by maximizing joint likelihood. Neurocomputing **22**, 49–67 (1998)

24. Jagannatham, A., Rao, B.: Cramér-Rao lower bound for constrained complex parameters. IEEE Signal Process. Lett. **11**(11), 875–878 (2004)

25. Joho, M., Mathis, H., Lambert, R.H.: Overdetermined blind source separation: using more sensors than source signals in a noisy mixture. In: Proceedings of the International Conference on Independent Component Analysis and Blind Source Separation (ICA), pp. 81–86 (2000)

26. Koldovsky, Z., Tichavsky, P.: Methods of fair comparison of performance of linear ICA techniques in presence of additive noise. In: Proceedings of the IEEE Conference on Acoustics, Speech, and Signal Processing (ICASSP), vol. 5, pp. 873–876 (2006)

27. Koldovsky, Z., Tichavsky, P.: Asymptotic analysis of bias of Fast ICA-based algorithms in the presence of additive noise. Technical Report 2181, UTIA, AV CR (2007a)

28. Koldovsky, Z., Tichavsky, P.: Blind instantaneous noisy mixture separation with best interference-plus-noise rejection. In: Proceedings of the International Conference on Independent Component Analysis and Blind Source Separation (ICA), pp. 730–737 (2007b)

29. Li, H., Adali, T.: Algorithms for complex ML ICA and their stability analysis using Wirtinger calculus. IEEE Trans. Signal Process. 58(12) (2010)

30. Li, X.L., Adali, T.: Complex independent component analysis by entropy bound minimization. IEEE Trans. Circ. Syst. I Regul. Pap. **57**(7), 1417–1430 (2010)

31. Loesch, B.: Complex blind source separation with audio applications. Ph.D. thesis, University of Stuttgart (2013). http://www.hut-verlag.de/9783843911214.html

32. Loesch, B., Yang, B.: On the relation between ICA and MMSE based source separation. In: Proceedings of the IEEE Conference on Acoustics, Speech, and Signal Processing (ICASSP), pp. 3720–3723 (2011)

33. Loesch, B., Yang, B.: Cramér-Rao bound for circular complex independent component analysis. In: Proceedings of the International Conference on Latent Variable Analysis and Signal Separation (LVA/ICA), pp. 42–49 (2012a)

34. Loesch, B., Yang, B.: On the solution of circular and noncircular complex KLD-ICA in the presence of noise. In: Proceedings of the European Signal Processing Conference (EUSIPCO), pp. 1479–1483 (2012b)

35. Loesch, B., Yang, B.: Cramér-Rao bound for circular and noncircular complex independent component analysis. IEEE Trans. Signal Process. **61**(2), 365–379 (2013)

36. Mandic, D.P., Goh, V.S.L.: Complex Valued Nonlinear Adaptive Filters: Noncircularity, Widely Linear, and Neural Models, 1st edn. Wiley, Chichester (2009)

37. Novey, M., Adali, T.: ICA by maximization of nongaussianity using complex functions. In: Proceedings of the IEEE Workshop on Machine Learning for Signal Processing (MLSP), pp. 21–26 (2005)
38. Novey, M., Adali, T.: Adaptable nonlinearity for complex maximization of nongaussianity and a fixed-point algorithm. In: Proceedings of the IEEE Workshop on Machine Learning for Signal Processing (MLSP), pp. 79–84 (2006)
39. Novey, M., Adali, T.: On extending the complex FastICA algorithm to noncircular sources. IEEE Trans. Signal Process. **56**(5), 2148–2154 (2008)
40. Novey, M., Adali, T., Roy, A.: A complex generalized gaussian distribution—characterization, generation, and estimation. IEEE Trans. Signal Process. **58**(3), 1427–1433 (2010)
41. Ollila, E., Kim, H.J., Koivunen, V.: Compact Cramér-Rao bound expression for independent component analysis. IEEE Trans. Signal Process. **56**(4), 1421–1428 (2008)
42. Ollila, E., Koivunen, V., Eriksson, J.: On the Cramér-Rao bound for the constrained and unconstrained complex parameters, pp. 414–418 (2008)
43. Remmert, R.: Theory of Complex Functions. Graduate Texts in Mathematics. Springer, New York (1991)
44. Schreier, P.J., Scharf, L.L.: Statistical signal processing of complex-valued data: The theory of improper and noncircular signals. Cambridge University Press, Cambridge (2010)
45. Tichavsky, P., Koldovsky, Z., Oja, E.: Performance analysis of the Fast ICA algorithm and Cramér-Rao bounds for linear independent component analysis. IEEE Trans. Signal Process. **54**(4) (2006)
46. Wirtinger, W.: Zur formalen theorie der funktionen von mehr komplexen veränderlichen. Math. Ann. **97**(1), 357–375 (1927)
47. Yeredor, A.: Blind separation of gaussian sources with general covariance structures: bounds and optimal estimation. IEEE Trans. Signal Process. **58**(10), 5057–5068 (2010)
48. Yeredor, A.: Performance analysis of the strong uncorrelating transformation in blind separation of complex-valued sources. IEEE Trans. Signal Process. **60**(1), 478–483 (2012)
49. Zhang, L.Q., Cichocki, A., Amari, S.: Natural gradient algorithm for blind separation of overdetermined mixture with additive noise. IEEE Signal Process. Lett. **6**(11), 293–295 (1999)
50. Zhu, X.L., Zhang, X.D., Ye, J.M.: A generalized contrast function and stability analysis for overdetermined blind separation of instantaneous mixtures. Neural Comput. **18**(3), 709–728 (2006)