ON THE SOLUTION OF CIRCULAR AND NONCIRCULAR COMPLEX KLD-ICA IN THE PRESENCE OF NOISE

Benedikt Loesch and Bin Yang

Institute of Signal Processing and System Theory, University of Stuttgart
Email: {benedikt.loesch,bin.yang}@ISS.uni-stuttgart.de

ABSTRACT

This paper aims at studying the solution of linear independent component analysis (ICA) based on Kullback-Leibler divergence (KLD) for a linear noisy mixing model in the determined case. The derivation is done using a perturbation analysis valid for small noise variance. We study the noncircular complex and the circular complex case. We show that for a wide range of both the shape parameter and the noncircularity index of the generalized Gaussian distribution (GGD), the signal-to-interference-plus-noise ratio (SINR) of KLD-based ICA is close to that of linear minimum mean squared error (MMSE) estimation.

Index Terms— Blind source separation, Independent component analysis, Minimum mean square error, Kullback-Leibler divergence, Perturbation analysis, noncircular complex

1. INTRODUCTION

In the last decades, independent component analysis (ICA) has been studied intensively and has found many applications [1]. Usually, a linear noiseless mixing model is assumed and the separated signals are obtained using a linear transform of the observed signals. Although noisy mixing models have been studied quite early, for example in the context of contrast-based approaches [1], many publications still consider a noiseless mixing model for simplicity. The presence of noise leads to a bias in the estimation of the mixing matrix. Douglas et. al. introduced measures in [2] to reduce this bias. Cardoso showed in [3] that for the noisy case the performance of source separation depends on the distribution of the sources, the signal-to-noise (SNR) ratio and the mixing matrix. Hyvärinen showed in [4] that in the noisy case, the maximum-likelihood (ML) estimate of the signals is a nonlinear function of the observations. Koldovsky et. al. derived in [5] the bias of several variants of the real FastICA algorithm from the MMSE solution. Davies studied in [6] identifiability issues in noisy real ICA.

In this paper, we focus on KLD-based ICA with equal number of sources and sensors (determined case). As shown by many different authors, KLD is (up to constant terms) identical to mutual information, the information maximization principle and for the noiseless case to ML estimation [1]. In many applications such as audio source separation in the frequency domain or telecommunication, signals are complex and not necessarily circular. General conditions regarding identifiability, uniqueness and separability can be found in [7]. Algorithms for complex ML-ICA have been studied for the noiseless case in [8]. However, practical applications have to deal with noise [9]. In [10], we derived the theoretical ICA solution for the noisy determined case and real signals. In this paper, we extend the results to the circular and noncircular complex case which is not directly equivalent to the real case of twice larger dimension [7].

In the estimation of the demixing matrix, one has to distinguish two different factors:

1. 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.
2. Variance of the estimated inverse mixing matrix in the noiseless case due to blind estimation and randomness of the sources. This variance can be lower bounded by the Cramér-Rao bound for ICA derived for the real case in [11] and for the circular complex case in [12].

In this paper, we study only the first factor, i.e. the bias of the demixing matrix from the inverse mixing matrix.

2. SIGNAL MODEL

We assume the linear noisy mixing model

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

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

1. The deterministic mixing matrix $\mathbf{A} \in \mathbb{C}^{N\times N}$ is invertible.
2. $\mathbf{s} = [s_1, \cdots, s_N]^T \in \mathbb{C}^N$ are N independent random variables with zero mean, unit variance and second-order noncircularity index $\gamma_i = E[s_i^2] \in [0, 1]$ (after scaling the columns of $\mathbf{A}$ suitably). Since $\gamma_i \in \mathbb{R}$, the real and imaginary part of s_i are uncorrelated. Then $\gamma_i \neq 0$ if and only if the variance of real and imaginary part differ.
   We further assume: The probability density functions (pdfs) $q_i(s_i)$ of s_i can be different. $q_i(s_i)$ is three times continuously differentiable with respect to s_i and s_i^* in the sense of Wirtinger derivatives which will be introduced in section 3.1. All expectations required in this paper exist.
3. $\mathbf{v} = [v_1, \cdots, v_N]^T \in \mathbb{C}^N$ are N random variables with zero mean and the covariance matrix $E[\mathbf{v}\mathbf{v}^H] = \sigma^2\mathbf{R}_\mathbf{v}$. $\sigma^2 = \frac{1}{N}\mathrm{tr}\left[E(\mathbf{v}\mathbf{v}^H)\right]$ is the average variance of $\mathbf{v}$ and $\mathrm{tr}(\mathbf{R}_\mathbf{v}) = N$. $\bar{\mathbf{R}}_\mathbf{v} = \frac{1}{\sigma^2}E[\mathbf{v}\mathbf{v}^T]$ 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. $q(\mathbf{v}) = q(-\mathbf{v})$. 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.
4. $\mathbf{s}$ and $\mathbf{v}$ are independent.

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

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

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

3. KLD-BASED ICA

In this paper, we focus on the ICA solution based on the KLD

$$D_{pq}(\mathbf{W}) = \int p_{\mathbf{y}}(\mathbf{y};\mathbf{W}) \log \frac{p_{\mathbf{y}}(\mathbf{y};\mathbf{W})}{q(\mathbf{y})} d\mathbf{y} \tag{3}$$

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 $\mathbf{W}$. $q(\mathbf{s}) = \prod_{i=1}^{N} q_i(s_i)$ is the assumed pdf of the original signals. We assume that we have perfect knowledge about the distribution of the original signals and $q(\mathbf{s})$ is identical to the true pdf $q^0(\mathbf{s})$ of $\mathbf{s}$. The KLD is known to have the following properties:

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

This means that 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 $q(\mathbf{s})$. Since we assume $q(\mathbf{s}) = \prod_{i=1}^{N} q_i(s_i)$, minizing KLD corresponds to making $\mathbf{y}$ as independent as possible and the individual y_i to have a pdf as close as possible to $q_i(s_i)$. The ICA solution $\mathbf{W}_{\text{ICA}}$ for the demixing matrix based on KLD is given by

$$\mathbf{W}_{\text{ICA}} = \arg\min_{\mathbf{W}} D_{pq}(\mathbf{W}). \tag{4}$$

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.

3.1. 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 [13], we can also write it as a function of $\mathbf{W}$ and $\mathbf{W}^*$:

$$D_{pq}(\mathbf{W},\mathbf{W}^*) = \int p_{\mathbf{y}}(\mathbf{y},\mathbf{y}^*;\mathbf{W},\mathbf{W}^*) \log \frac{p_{\mathbf{y}}(\mathbf{y},\mathbf{y}^*;\mathbf{W},\mathbf{W}^*)}{q(\mathbf{y},\mathbf{y}^*)} d\mathbf{y}. \tag{5}$$

Before we continue with the derivation of the ICA solution, we provide a short summary of the Wirtinger calculus. Given $\mathbf{Z} = \mathbf{X} + j\mathbf{Y} \in \mathbb{C}^{N\times M}$, $\mathbf{X},\mathbf{Y} \in \mathbb{R}^{N\times M}$, and a real scalar cost function $f(\mathbf{Z},\mathbf{Z}^*) = \tilde{f}(\mathbf{X},\mathbf{Y}) \in \mathbb{R}$. Instead of calculating the derivatives of $\tilde{f}(\cdot)$ with respect to $\mathbf{X},\mathbf{Y}$, the Wirtinger calculus calculates the partial derivatives of $f(\mathbf{Z},\mathbf{Z}^*)$ with respect to $\mathbf{Z}$ and $\mathbf{Z}^*$, treating $\mathbf{Z}$ and $\mathbf{Z}^*$ as two independent variables [13]. Let $\frac{\partial f}{\partial \mathbf{Z}}, \frac{\partial f}{\partial \mathbf{Z}^*} \in \mathbb{C}^{N\times M}$ and $\frac{\partial \tilde{f}}{\partial \mathbf{X}}, \frac{\partial \tilde{f}}{\partial \mathbf{Y}} \in \mathbb{R}^{N\times M}$. It can be shown:

1. The partial derivatives of $f(\cdot)$ with respect to $\mathbf{Z}$ and $\mathbf{Z}^*$ are
$$\frac{\partial f}{\partial \mathbf{Z}} = \frac{1}{2}\left(\frac{\partial \tilde{f}}{\partial \mathbf{X}} - j\frac{\partial \tilde{f}}{\partial \mathbf{Y}}\right), \quad \frac{\partial f}{\partial \mathbf{Z}^*} = \frac{1}{2}\left(\frac{\partial \tilde{f}}{\partial \mathbf{X}} + j\frac{\partial \tilde{f}}{\partial \mathbf{Y}}\right). \tag{6}$$
2. The stationary point of $f(\cdot)$ and $\tilde{f}(\cdot)$ is given by
$$\frac{\partial \tilde{f}}{\partial \mathbf{X}} = \mathbf{0} \text{ and } \frac{\partial \tilde{f}}{\partial \mathbf{Y}} = \mathbf{0} \Leftrightarrow \frac{\partial f}{\partial \mathbf{Z}} = \mathbf{0} \Leftrightarrow \frac{\partial f}{\partial \mathbf{Z}^*} = \mathbf{0}. \tag{7}$$
3. The direction of steepest descent of $f(\cdot)$ is given by $-\frac{\partial f}{\partial \mathbf{Z}^*}$ and not $-\frac{\partial f}{\partial \mathbf{Z}}$.

Coming back to the KLD cost function in (5), its derivative $\frac{\partial D_{pq}(\mathbf{W},\mathbf{W}^*)}{\partial \mathbf{W}^*}$ is given by [8]

$$\frac{\partial D_{pq}(\mathbf{W},\mathbf{W}^*)}{\partial \mathbf{W}^*} = -\mathbf{W}^{-H} + E\left[\boldsymbol{\varphi}^*(\mathbf{y},\mathbf{y}^*)\mathbf{x}^H\right], \tag{8}$$

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

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

$$\left.\frac{\partial D_{pq}(\mathbf{W},\mathbf{W}^*)}{\partial \mathbf{W}^*}\right|_{\mathbf{W}=\mathbf{W}_{\text{ICA}}} \stackrel{!}{=} \mathbf{0} \quad \text{or}$$

$$E(\boldsymbol{\varphi}^*(\mathbf{y}_{\text{ICA}},\mathbf{y}_{\text{ICA}}^*)\mathbf{y}_{\text{ICA}}^H) = E(\boldsymbol{\varphi}(\mathbf{y}_{\text{ICA}},\mathbf{y}_{\text{ICA}}^*)\mathbf{y}_{\text{ICA}}^T)^* \stackrel{!}{=} \mathbf{I} \tag{9}$$

with $\mathbf{y}_{\text{ICA}} = \mathbf{W}_{\text{ICA}}\mathbf{x} = \mathbf{W}_{\text{ICA}}\mathbf{A}\mathbf{s} + \mathbf{W}_{\text{ICA}}\mathbf{v} = \hat{\mathbf{y}} + \mathbf{W}_{\text{ICA}}\mathbf{v}$. The properties of the ICA solution are:

- $\mathbf{W}_{\text{ICA}} = \mathbf{A}^{-1}$ if $\sigma^2 = 0$ (no noise) and $q(\mathbf{s}) = q^0(\mathbf{s})$.
- To compute $\mathbf{W}_{\text{ICA}}$, we do not need to know $\mathbf{A}$ or $\mathbf{s}$, but the pdf $q(\mathbf{s}) = \prod_{i=1}^{N} q_i(s_i)$ is required. All $q_i(s_i)$ must either be non-Gaussian or Gaussian with distinct noncircularity indices.
- No permutation ambiguity if $q_i(s_i) \neq q_j(s_j)\ \forall i \neq j$.
- There is no scaling ambiguity if $q_i(s_i) = q_i^0(s_i)$ is known $\forall i$. Only a phase ambiguity remains if $q_i(s_i)$ is circular.

As shown in the appendix, 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

$$\mathbf{W}_{\text{ICA}} = (\mathbf{I} + \sigma^2\mathbf{C})\mathbf{A}^{-1} + O(\sigma^4). \tag{10}$$

The elements of $\mathbf{C}$ can be obtained from Eq. (33) and (35). If $q(\mathbf{s})$ is symmetric in the real or imaginary part of $\mathbf{s}$, they are given by Eq. (34) and (36).

For comparison, we consider the linear MMSE estimator

$$\mathbf{W}_{\text{MMSE}} = \mathbf{A}^H\left(\mathbf{A}\mathbf{A}^H + \sigma^2\mathbf{R}_{\mathbf{v}}\right)^{-1} \tag{11}$$

$$= \left[\mathbf{I} - \sigma^2\mathbf{R}_{-1}\right]\mathbf{A}^{-1} + O(\sigma^4). \tag{12}$$

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 (12) with (10) we see that $\mathbf{W}_{\text{ICA}}$ and $\mathbf{W}_{\text{MMSE}}$ are similar if $\mathbf{C} \approx -\mathbf{R}_{-1}$.

3.2. Circular complex case

A complex random vector $\mathbf{s}$ is circular if $\mathbf{s}$ and $\mathbf{s}e^{j\theta}$ have the same pdf for any θ [13]. 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, (34) and (36) simplify to

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

$$C_{ij} = -\frac{\kappa_j(\kappa_i - 1)}{\kappa_i\kappa_j - 1}[\mathbf{R}_{-1}]_{ij} \in \mathbb{C} \quad (i \neq j). \tag{13}$$

3.3. Real case

For real signals and noise, we have

$$\gamma_i = 1, \quad \mathbf{R}_{\mathbf{v}} = \bar{\mathbf{R}}_{\mathbf{v}}. \tag{14}$$

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 (6)) and hence

$$\xi_i = \kappa_i, \quad \rho_i = \delta_i, \quad \lambda_i = \omega_i = \tau_i. \tag{15}$$

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 (6)). Hence it holds

$$\kappa_i = \frac{\tilde{\kappa}_i}{2}, \quad \rho_i = \frac{\tilde{\rho}_i}{2}, \quad \lambda_i = \frac{\tilde{\lambda}_i}{4}, \tag{16}$$

where $\tilde{\kappa}_i$, $\tilde{\rho}_i$ and $\tilde{\lambda}_i$ are defined using real derivatives of $\varphi(s)$. Using (14)-(16), we get from (34) and (36)

$$C_{ii} = -\frac{\tilde{\kappa}_i + \frac{1}{2}\tilde{\lambda}_i}{1+\tilde{\rho}_i}\,[\mathbf{R}_{-1}]_{ii}\,,$$
$$C_{ij} = -\frac{\tilde{\kappa}_j(\tilde{\kappa}_i - 1)}{\tilde{\kappa}_i\tilde{\kappa}_j - 1}\,[\mathbf{R}_{-1}]_{ij} \quad (i \neq j). \tag{17}$$

which corresponds to the results in [10].

4. RESULTS FOR COMPLEX GGD

Since the real case has already been studied in [10], we focus on the complex case here. Due to space limitations, we study KLD-ICA only 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 in [5], we use the signal-to-interference-plus-noise ratio (SINR) to evaluate separation performance. For spatially uncorrelated noise, SINR for a given demixing matrix $\mathbf{W}$ is calculated as

$$\text{SINR} = \frac{1}{N}\sum_{i=1}^{N}\frac{|[\mathbf{WA}]_{ii}|^2}{\sum_{j\neq i}|[\mathbf{WA}]_{ij}|^2 + \sigma^2\sum_j|\mathbf{W}_{ij}|^2}. \tag{18}$$

The term $|[\mathbf{WA}]_{ii}|^2$ reflects the power of the desired source i in the demixed signal y_i. $\sum_{j\neq i}|[\mathbf{WA}]_{ij}|^2$ corresponds to the power of the interfering signals $j \neq i$ in the demixed signal y_i. $\sigma^2\sum_j|\mathbf{W}_{ij}|^2$ is the noise power in the demixed signal y_i. The signal-to-noise ratio is defined as $1/\sigma^2$, i.e. it denotes the SNR before mixing. It is known that among all linear demixing matrices $\mathbf{W}$, $\mathbf{W}_{\text{MMSE}}$ from (11) is the one which maximizes the SINR. We compare the SINR of the theoretical ICA solution $\mathbf{W}_{\text{ICA}}$ from (10), the average SINR of $\hat{\mathbf{W}}_{\text{ICA}}$ obtained from 100 runs of KLD-based ICA using L samples and the SINR of $\mathbf{W}_{\text{MMSE}}$ from (11). The ICA algorithm is initialized with $\mathbf{W} = \mathbf{I}$ and performs gradient descent using the relative gradient [1], i.e. postmultiplies the gradient of KLD (8) 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 $\mathbf{W}_{\text{ICA}}$, we evaluated all expectations exactly. Hence $\mathbf{W}_{\text{ICA}}$ only accounts for the bias from $\mathbf{A}^{-1}$ but not for estimation variance whereas $\hat{\mathbf{W}}_{\text{ICA}}$ contains both factors.

In the following, all sources are GGD with the same shape parameter $c_i = c$. The pdf of a noncircular complex GGD with zero mean and $E[|s|^2] = 1$ can be written as [14] $q(s,s^*) = \frac{c\alpha}{\pi\Gamma(1/c)(1-\gamma^2)^{1/2}}\exp\left(-\left[\frac{\alpha/2}{\gamma^2-1}\left(\gamma s^2 + \gamma s^{*2} - 2ss^*\right)\right]^c\right)$, where the scale parameter $\alpha = \Gamma(2/c)/\Gamma(1/c)$ and $\Gamma(\cdot)$ is the Gamma function. $\gamma \in [0,1]$ controls the noncircularity, i.e. imbalance between the power of real and imaginary part. A change of 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. By integration in polar coordinates, it can be shown that

$$\kappa = E[\theta(s)] = \int\frac{\partial\varphi}{\partial s^*}q^0(s)ds = \frac{c^2\Gamma(2/c)}{(1-\gamma^2)\Gamma^2(1/c)}, \tag{19}$$
$$\delta = E[\theta(s)s^*s] = \int\frac{\partial\varphi}{\partial s^*}ss^*q^0(s) = \frac{2c+(1-c)\gamma^2}{2(1-\gamma^2)}, \tag{20}$$
$$\rho = E[\eta(s)s^2] = \int\frac{\partial\varphi}{\partial s}s^2q^0(s)ds = -\frac{2c-2+(1-3c)\gamma^2}{2(1-\gamma^2)}, \tag{21}$$
$$\xi = E[\eta(s)] = \int\frac{\partial\varphi}{\partial s}q^0(s)ds = -\gamma\kappa, \tag{22}$$
$$\lambda = E[\epsilon(s)s] = \int\frac{\partial^2\varphi}{\partial s\partial s^*}sq^0(s)ds = (c-1)\kappa, \tag{23}$$
$$\omega = E[\nu(s)s] = \int\frac{\partial^2\varphi}{(\partial s)^2}sq^0(s)ds = -\frac{3}{2}(c-1)\gamma\kappa, \tag{24}$$
$$\tau = E[\zeta(s)s] = \int\frac{\partial^2\varphi}{(\partial s^*)^2}sq^0(s)ds = -\frac{1}{2}(c-1)\gamma\kappa. \tag{25}$$

4.1. 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 $\xi = \omega = \tau = 0$. Fig. 1 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 Cramér-Rao bound approaches infinity [12]. Hence estimation of $\mathbf{W}$ becomes impossible. This is reflected in Fig. 1: The SINR for $\hat{\mathbf{W}}_{\text{ICA}}$ estimated by KLD-ICA decreases for $c \to 1$.

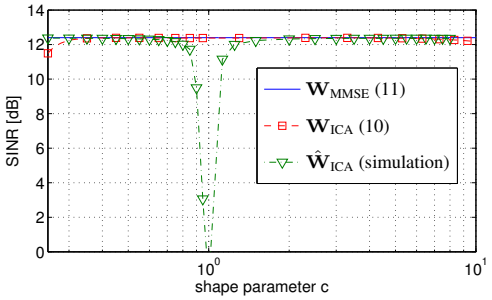
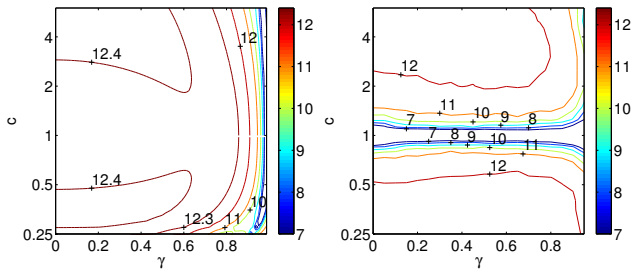


Fig. 1: SINR for circular complex GGD signal and circular complex noise, SNR $= 10$ dB, $\hat{\mathbf{W}}_{\text{ICA}}$ uses $L = 10^4$ samples.

4.2. 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. Fig. 2 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. 2 (a) and (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 $c \to 1$. Hence, in the simulation the SINR decreases in the vicinity of $c = 1$. Furthermore, the smaller sample size of $L = 10^3$ leads to a larger variance of $\hat{\mathbf{W}}_{\text{ICA}}$ which is not reflected in the theoretical ICA solution $\mathbf{W}_{\text{ICA}}$. However, Fig. 2 (b) shows that even with a limited sample size $\hat{\mathbf{W}}_{\text{ICA}}$ estimated by KLD-ICA can still achieve an SINR quite close to that of MMSE except for $c \approx 1$.

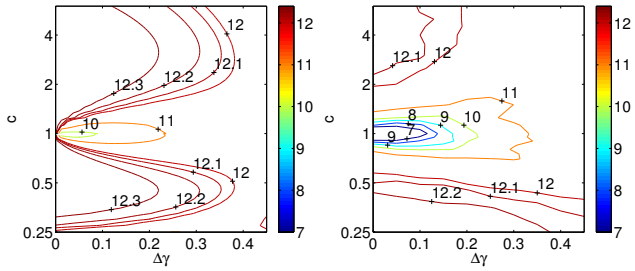
Now, we consider the case where sources are noncircular complex with $\gamma_1 = 0.5, \gamma_{2,3} = 0.5 \pm \Delta\gamma$ and the noise $\mathbf{v}$ is noncircular with $\mathbf{R}_{\mathbf{v}} = \mathbf{I}$ and $\bar{\mathbf{R}}_{\mathbf{v}} = 0.5\cdot\mathbf{I}$, i.e. $\gamma_{\text{noise}} = 0.5$. Fig. 3 shows decreasing SINR values for $c \to 1$ and $\Delta\gamma \to 0$ since in that region $|\Re C_{ij}|$ in (36) becomes large if sources or noise are noncircular. However, except for this region, the SINR of the theoretical ICA solution (Fig. 3 (a)) is still close to that of MMSE. The form of the contour plot for the simulation (Fig. 3 (b)) is similar to that of the theoretical solu-



(a) $\mathbf{W}_{\text{ICA}}$ (10) (b) $\hat{\mathbf{W}}_{\text{ICA}}$ (simulation, $L = 10^3$)

Fig. 2: SINR [dB] of ICA solution for noncircular complex GGD signals with $\gamma_i = \gamma$, circular complex noise and SNR $= 10\,\text{dB}$

tion 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 $\Delta\gamma$. 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.



(a) $\mathbf{W}_{\text{ICA}}$ (10) (b) $\hat{\mathbf{W}}_{\text{ICA}}$ (simulation, $L = 10^3$)

Fig. 3: 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 $= 10\,\text{dB}$

In summary, the results in Sect. 4.1 and 4.2 have shown that

- in many cases the theoretical solution $\mathbf{W}_{\text{ICA}}$ of KLD-ICA can achieve an SINR close to the optimum attainable by any linear demixing matrix $\mathbf{W}$ which is given by $\mathbf{W} = \mathbf{W}_{\text{MMSE}}$
- for sources following a GGD, $\hat{\mathbf{W}}_{\text{ICA}}$ obtained by running KLD-ICA with a finite amount of samples L can achieve an SINR quite close to that of $\mathbf{W}_{\text{MMSE}}$ except for (nearly) Gaussian sources with similar noncircularity indices.

Although we assumed that we perfectly know the distributions of the sources, other approaches such as ICA-EBM [15] exist which do not require such knowledge. Simulation results using ICA-EBM show similar SINR performance as KLD-ICA.

5. CONCLUSION AND FUTURE WORK

We have derived an analytic expression for the demixing matrix of KLD-based ICA for the low noise regime by a perturbation analysis taking into account all terms of order σ^2. In this paper, we have considered the general noncircular complex determined case. The solution for the circular complex and real case can be derived as special cases. 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 paper but also the variance of the estimate are important for SINR. Our derivation for the low noise regime could be extended for large noise by including higher-order terms in the Taylor series.

APPENDIX A. DERIVATION OF (10)

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

1. Taylor series approximation of $E(\boldsymbol{\varphi}(\mathbf{y})\mathbf{y}^T)$ in (9) at $\mathbf{y} = \hat{\mathbf{y}} = \mathbf{W}_{\text{ICA}}\mathbf{A}\mathbf{s}$ taking into account all terms of order σ^2.
2. Taylor series approximation of the result of the above step by exploiting $\mathbf{W}_{\text{ICA}} = \mathbf{A}^{-1} + \sigma^2\mathbf{B} + O(\sigma^4)$ and $\hat{\mathbf{y}} = \mathbf{s} + \sigma^2\mathbf{B}\mathbf{A}\mathbf{s} + O(\sigma^4) = \mathbf{s} + \sigma^2\mathbf{C}\mathbf{s} + O(\sigma^4) = \mathbf{s} + \sigma^2\mathbf{b} + O(\sigma^4)$ with $\mathbf{C} = \mathbf{B}\mathbf{A}$ and $\mathbf{b} = \mathbf{C}\mathbf{s} = [b_1, \cdots, b_N]^T$.

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

The general Taylor series expansion of $\varphi(y) \mathrel{\widehat{=}} \varphi(y, y^*)$ is given in (26) on the next page with $\eta(y, y^*) = \frac{\partial\varphi}{\partial y}$, $\theta(y, y^*) = \frac{\partial\varphi}{\partial y^*}$, $\nu(y, y^*) = \frac{\partial^2\varphi}{(\partial y)^2}$, $\zeta(y, y^*) = \frac{\partial^2\varphi}{(\partial y^*)^2}$ and $\epsilon(y, y^*) = \frac{\partial^2\varphi}{\partial y \partial y^*}$. To simplify notation, we will drop the dependence of $\varphi(\cdot), \eta(\cdot), \theta(\cdot), \nu(\cdot), \zeta(\cdot), \epsilon(\cdot)$ on y^* and keep only the dependence on y in the following. Furthermore, boldface vectors $\mathbf{z}$ denote the corresponding scalar values $z_i, 1 \le i \le N$ stacked into an N-dimensional column vector.

For the first Taylor series, we expand $\boldsymbol{\varphi}(\mathbf{y})$ at $\mathbf{y} = \hat{\mathbf{y}} = \mathbf{W}\mathbf{A}\mathbf{s}$. By using a vectorized form of (26) with $\boldsymbol{\Delta}\mathbf{y} = \mathbf{W}\mathbf{v}$ we get (27) on the next page. $\text{diag}(\mathbf{z})$ is a diagonal matrix whose diagonal elements are those of the vector $\mathbf{z}$. $\text{Diag}(\mathbf{Z})$ sets all off-diagonal elements of the matrix $\mathbf{Z}$ to zero. Multiplying (27) by $\mathbf{y}^T$ and taking the expectation, we get (28) from (9).

The (i, j)-th element of the matrix $E\left[\boldsymbol{\varphi}(\hat{\mathbf{y}})\hat{\mathbf{y}}^T\right]$ in (28) can be written as a second Taylor series developed at $\hat{\mathbf{y}} = \mathbf{s}$:

$$
\begin{aligned}
&E\left[\boldsymbol{\varphi}(\hat{\mathbf{y}})\hat{\mathbf{y}}^T\right]_{ij} \\
&= E\left[(\varphi_i(s_i) + \eta_i(s_i)\sigma^2 b_i + \theta_i(s_i)\sigma^2 b_i^*)(s_j + \sigma^2 b_j)\right] + O(\sigma^4) \\
&= E\left[\varphi_i(s_i)s_j\right] + E\left[\eta_i(s_i)\sigma^2 b_i s_j\right] + E\left[\theta_i(s_i)\sigma^2 b_i^* s_j\right] \\
&\quad + E\left[\varphi_i(s_i)\sigma^2 b_j\right] + O(\sigma^4) \qquad (30)
\end{aligned}
$$

with $b_j = \sum_{l=1}^N C_{jl}s_l$. We calculate the individual terms as

$$
\begin{aligned}
E\left[\varphi_i(s_i)s_j\right] &= \begin{cases} 1 & i = j \\ 0 & i \ne j \end{cases} \\
E\left[\eta_i(s_i)b_i s_j\right] &= \sum_l C_{il}E\left[\eta_i(s_i)s_l s_j\right] = \begin{cases} \rho_i C_{ii} & i = j \\ \gamma_j\xi_i C_{ij} & i \ne j \end{cases} \\
E\left[\theta_i(s_i)b_i^* s_j\right] &= \sum_l C_{il}^* E\left[\theta_i(s_i)s_l^* s_j\right] = \begin{cases} \delta_i C_{ii}^* & i = j \\ \kappa_i C_{ij}^* & i \ne j \end{cases} \\
E\left[\varphi_i(s_i)b_j\right] &= C_{ji} \qquad (31)
\end{aligned}
$$

with $\rho_i = E\left[\eta_i(s_i)s_i^2\right]$, $\delta_i = E\left[\theta_i(s_i)s_i^* s_i\right]$, $\kappa_i = E\left[\theta_i(s_i)\right]$, $\xi_i = E\left[\eta_i(s_i)\right]$ and $\gamma_i = E[s_i^2]$. γ_i is the noncircularity index for source i. From (30), we get

$$
\begin{aligned}
E\left[\boldsymbol{\varphi}(\hat{\mathbf{y}})\hat{\mathbf{y}}^T\right] &= \mathbf{I} + \sigma^2(\text{diag}(\boldsymbol{\rho}) - \text{Diag}(\boldsymbol{\xi}\boldsymbol{\gamma}^T))\text{Diag}(\mathbf{C}) \\
&\quad + \sigma^2\text{diag}(\boldsymbol{\xi})\mathbf{C}\text{diag}(\boldsymbol{\gamma}) + \sigma^2\text{diag}(\boldsymbol{\kappa})\mathbf{C}^* \\
&\quad + \sigma^2\text{diag}(\boldsymbol{\delta} - \boldsymbol{\kappa})\text{Diag}(\mathbf{C}^*) + \sigma^2\mathbf{C}^T. \qquad (32)
\end{aligned}
$$

$$\varphi(y,y^*) = \varphi(\hat{y},\hat{y}^*) + \frac{\partial\varphi}{\partial y}\Delta y + \frac{\partial\varphi}{\partial y^*}\Delta y^* + \frac{1}{2}\left(\frac{\partial^2\varphi}{(\partial y)^2}(\Delta y)^2 + \frac{\partial^2\varphi}{(\partial y^*)^2}(\Delta y^*)^2\right) + \frac{\partial^2\varphi}{\partial y\partial y^*}\Delta y\Delta y^* + \dots$$
$$= \varphi(\hat{y},\hat{y}^*) + \eta(y,y^*)\Delta y + \theta(y,y^*)\Delta y^* + \frac{1}{2}\left(\nu(y,y^*)(\Delta y)^2 + \zeta(y,y^*)(\Delta y^*)^2\right) + \epsilon(y,y^*)\Delta y\Delta y^* + \dots \quad (26)$$

$$\boldsymbol{\varphi}(\mathbf{y}) = \boldsymbol{\varphi}(\hat{\mathbf{y}}) + \mathrm{diag}(\boldsymbol{\eta}(\hat{\mathbf{y}}))\mathbf{W}\mathbf{v} + \mathrm{diag}(\boldsymbol{\theta}(\hat{\mathbf{y}}))\mathbf{W}^*\mathbf{v}^* + \frac{1}{2}\mathrm{diag}(\boldsymbol{\nu}(\hat{\mathbf{y}}))(\mathbf{W}\mathbf{v}\odot\mathbf{W}\mathbf{v})$$
$$+ \frac{1}{2}\mathrm{diag}(\boldsymbol{\zeta}(\hat{\mathbf{y}}))(\mathbf{W}^*\mathbf{v}^*\odot\mathbf{W}^*\mathbf{v}^*) + \mathrm{diag}(\boldsymbol{\epsilon}(\hat{\mathbf{y}}))\mathbf{W}\mathbf{v}\odot\mathbf{W}^*\mathbf{v}^* + \dots \quad (27)$$

$$\mathbf{I} = E\left[\boldsymbol{\varphi}(\hat{\mathbf{y}})\hat{\mathbf{y}}^T\right] + E\left[\mathrm{diag}(\boldsymbol{\eta}(\hat{\mathbf{y}}))\right]E\left[\mathbf{W}\mathbf{v}\mathbf{v}^T\mathbf{W}^T\right] + E\left[\mathrm{diag}(\boldsymbol{\theta}(\hat{\mathbf{y}}))\right]E\left[\mathbf{W}^*\mathbf{v}^*\mathbf{v}^T\mathbf{W}^T\right] + \frac{1}{2}E\left[\boldsymbol{\nu}(\hat{\mathbf{y}})\hat{\mathbf{y}}^T\right]E\left[\mathrm{Diag}(\mathbf{W}\mathbf{v}\mathbf{v}^T\mathbf{W}^T)\right]$$
$$+ \frac{1}{2}E\left[\boldsymbol{\zeta}(\hat{\mathbf{y}})\hat{\mathbf{y}}^T\right]E\left[\mathrm{Diag}(\mathbf{W}^*\mathbf{v}^*(\mathbf{v}^*)^T(\mathbf{W}^*)^T)\right] + E\left[\boldsymbol{\epsilon}(\hat{\mathbf{y}})\hat{\mathbf{y}}^T\right]E\left[\mathrm{Diag}(\mathbf{W}\mathbf{v}\mathbf{v}^H\mathbf{W}^H)\right] + O(\sigma^4) \quad (28)$$

$$\mathbf{0} = (\mathrm{diag}(\boldsymbol{\rho}) - \mathrm{Diag}(\boldsymbol{\xi}\boldsymbol{\gamma}^T))\mathrm{Diag}(\mathbf{C}) + \mathrm{diag}(\boldsymbol{\xi})\mathbf{C}\mathrm{diag}(\boldsymbol{\gamma}) + \mathrm{diag}(\boldsymbol{\kappa})\mathbf{C}^* + \mathrm{diag}(\boldsymbol{\delta}-\boldsymbol{\kappa})\mathrm{Diag}(\mathbf{C}^*) + \mathbf{C}^T + \mathrm{diag}(\boldsymbol{\kappa})(\mathbf{W}\mathbf{R}_{\mathbf{v}}\mathbf{W}^H)^*$$
$$+ \mathrm{diag}(\boldsymbol{\lambda})\mathrm{Diag}(\mathbf{W}\mathbf{R}_{\mathbf{v}}\mathbf{W}^H) + \mathrm{diag}(\boldsymbol{\xi})(\mathbf{W}\bar{\mathbf{R}}_{\mathbf{v}}\mathbf{W}^T) + \frac{1}{2}\mathrm{diag}(\boldsymbol{\omega})\mathrm{Diag}(\mathbf{W}\bar{\mathbf{R}}_{\mathbf{v}}\mathbf{W}^T) + \frac{1}{2}\mathrm{diag}(\boldsymbol{\tau})\mathrm{Diag}(\mathbf{W}\bar{\mathbf{R}}_{\mathbf{v}}\mathbf{W}^T)^* + O(\sigma^4) \quad (29)$$

Let $\omega_i = E[\nu_i(s_i)s_i]$, $\tau_i = E[\zeta_i(s_i)s_i]$ and $\lambda_i = E[\epsilon_i(s_i)s_i]$. For the remaining terms in (28), we use the following properties:

- For $i \neq j$, $E[\nu_i(\hat{y}_i)\hat{y}_j]$, $E[\zeta_i(\hat{y}_i)\hat{y}_j]$ and $E[\epsilon_i(\hat{y}_i)\hat{y}_j]$ are all $O(\sigma^2)$ since $\hat{y}_i = s_i + O(\sigma^2)$, $\hat{y}_j = s_j + O(\sigma^2)$ and s_i and s_j are independent.
- With the same reasoning, we get $E[\eta_i(\hat{y}_i)] = \xi_i + O(\sigma^2)$, $E[\theta_i(\hat{y}_i)] = \kappa_i + O(\sigma^2)$, $E[\nu_i(\hat{y}_i)\hat{y}_i] = \omega_i + O(\sigma^2)$, $E[\zeta_i(\hat{y}_i)\hat{y}_i] = \tau_i + O(\sigma^2)$ and $E[\epsilon_i(\hat{y}_i)\hat{y}_i] = \lambda_i + O(\sigma^2)$.

As defined in Sec. 2, $E[\mathbf{v}\mathbf{v}^H] = \sigma^2\mathbf{R}_{\mathbf{v}}$ and $E[\mathbf{v}\mathbf{v}^T] = \sigma^2\bar{\mathbf{R}}_{\mathbf{v}}$. Using these properties and (32), we get (29) from (28).

Since $\mathbf{W} = \mathbf{A}^{-1} + O(\sigma^2)$, we define the transformed noise covariance matrix $\mathbf{R}_{-1} = \mathbf{W}\mathbf{R}_{\mathbf{v}}\mathbf{W}^H = \mathbf{A}^{-1}\mathbf{R}_{\mathbf{v}}\mathbf{A}^{-H} + O(\sigma^2)$ and the transformed noise pseudo-covariance matrix $\bar{\mathbf{R}}_{-1} = \mathbf{W}\bar{\mathbf{R}}_{\mathbf{v}}\mathbf{W}^T = \mathbf{A}^{-1}\bar{\mathbf{R}}_{\mathbf{v}}\mathbf{A}^{-T} + O(\sigma^2)$. Note that $\mathbf{R}_{-1}^H = \mathbf{R}_{-1}$ and $\bar{\mathbf{R}}_{-1}^T = \bar{\mathbf{R}}_{-1}$. For the diagonal elements C_{ii} we get from (29)

$$\rho_i C_{ii} + \delta_i C_{ii}^* + C_{ii} = -(\kappa_i + \lambda_i)[\mathbf{R}_{-1}]_{ii} - (\xi_i + \frac{1}{2}\omega_i)[\bar{\mathbf{R}}_{-1}]_{ii} - \frac{1}{2}\tau_i[\bar{\mathbf{R}}_{-1}]_{ii}^*. \quad (33)$$

If $q(s,s^*)$ is symmetric in the real part $\Re s$ or imaginary part $\Im s$ of s, i.e. $q(-\Re s,\Im s) = q(\Re s,\Im s)$ or $q(\Re s,-\Im s) = q(\Re s,\Im s)$, 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

$$\Re C_{ii} = -\frac{(\kappa_i+\lambda_i)[\mathbf{R}_{-1}]_{ii} + (\xi_i + \frac{1}{2}(\omega_i+\tau_i))[\Re\bar{\mathbf{R}}_{-1}]_{ii}}{\rho_i + 1 + \delta_i},$$
$$\Im C_{ii} = -\frac{(\xi_i + \frac{1}{2}(\omega_i - \tau_i))[\Im\bar{\mathbf{R}}_{-1}]_{ii}}{\rho_i + 1 - \delta_i}. \quad (34)$$

For the off-diagonal elements C_{ij} we get from (29)

$$\gamma_j\xi_i C_{ij} + \kappa_i C_{ij}^* + C_{ji} = -\kappa_i[\mathbf{R}_{-1}]_{ij}^* - \xi_i[\bar{\mathbf{R}}_{-1}]_{ij},$$
$$\gamma_i\xi_j C_{ji} + \kappa_j C_{ji}^* + C_{ij} = -\kappa_j[\mathbf{R}_{-1}]_{ji}^* - \xi_j[\bar{\mathbf{R}}_{-1}]_{ji}. \quad (35)$$

If $q(s,s^*)$ is symmetric in the real or imaginary part of s, and if $(\gamma_j\xi_i+\kappa_i)(\gamma_i\xi_j+\kappa_j) \neq 1$ and $(\gamma_j\xi_i-\kappa_i)(\gamma_i\xi_j-\kappa_j) \neq 1$, we obtain the real and imaginary part of C_{ij} from (35) as

$$\Re C_{ij} = \frac{(\kappa_j - \kappa_i(\gamma_i\xi_j+\kappa_j))[\Re\mathbf{R}_{-1}]_{ij} + (\xi_j - \xi_i(\gamma_i\xi_j+\kappa_j))[\Re\bar{\mathbf{R}}_{-1}]_{ij}}{(\gamma_j\xi_i+\kappa_i)(\gamma_i\xi_j+\kappa_j)-1},$$
$$\Im C_{ij} = \frac{(\kappa_j + \kappa_i(\gamma_i\xi_j-\kappa_j))[\Im\mathbf{R}_{-1}]_{ij} + (\xi_j - \xi_i(\gamma_i\xi_j-\kappa_j))[\Im\bar{\mathbf{R}}_{-1}]_{ij}}{(\gamma_j\xi_i-\kappa_i)(\gamma_i\xi_j-\kappa_j)-1}. \quad (36)$$

REFERENCES

[1] Pierre Comon and Christian Jutten, Eds., *Handbook of blind source separation : independent component analysis and applications*, Elsevier, Amsterdam, 1st. ed. edition, 2010.

[2] S.C. Douglas, A. Cichocki, and S. Amari, "A bias removal technique for blind source separation with noisy measurements," *Eletronics Letters*, vol. 34, pp. 1379–1380, July 1998.

[3] Jean-Francois Cardoso, "On the performance of orthogonal source separation algorithms," in *Proc. EUSIPCO*, 1994.

[4] A. Hyvärinen, "Independent component analysis in the presence of Gaussian noise by maximizing joint likelihood," *Neurocomputing*, vol. 22, pp. 49–67, 1998.

[5] Z. Koldovsky and P. Tichavsky, "Blind instantaneous noisy mixture separation with best interference-plus-noise rejection," *Proc. ICA 2007*, Sept. 2007.

[6] M. Davies, "Identifiability issues in noisy ICA," *IEEE Signal Processing Letters*, vol. 11, no. 5, pp. 470–473, May 2004.

[7] J. Eriksson and V. Koivunen, "Complex random vectors and ICA models: identifiability, uniqueness, and separability," *Information Theory, IEEE Transactions on*, vol. 52, no. 3, pp. 1017 –1029, March 2006.

[8] H. Li and T. Adali, "Algorithms for complex ML ICA and their stability analysis using Wirtinger calculus," *IEEE Trans. on Signal Processing*, vol. 58, no. 12, Dec. 2010.

[9] S. Javidi, D.P. Mandic, and A. Cichocki, "Complex blind source extraction from noisy mixtures using second-order statistics," *Circuits and Systems I: Regular Papers, IEEE Transactions on*, vol. 57, no. 7, pp. 1404–1416, july 2010.

[10] B. Loesch and B. Yang, "On the relation between ICA and MMSE based source separation," *Proc. ICASSP*, 2011.

[11] P. Tichavsky, Z. Koldovsky, and E. Oja, "Performance analysis of the FastICA algorithm and Cramér-Rao bounds for linear independent component analysis," *IEEE Trans. on Signal Processing*, vol. 54, no. 4, Apr. 2006.

[12] B. Loesch and B. Yang, "Cramér-Rao bound for circular complex independent component analysis," *Proc. LVA/ICA*, 2012.

[13] T. Adali, P.J. Schreier, and L.L. Scharf, "Complex-valued signal processing: The proper way to deal with impropriety," *Signal Processing, IEEE Transactions on*, vol. 59, no. 11, pp. 5101 –5125, nov. 2011.

[14] M. Novey, T. Adali, and A. Roy, "A complex generalized gaussian distribution – characterization, generation, and estimation," *Signal Processing, IEEE Transactions on*, vol. 58, no. 3, pp. 1427 –1433, March 2010.

[15] Xi-Lin Li and T. Adali and, "Complex independent component analysis by entropy bound minimization," *Circuits and Systems I: Regular Papers, IEEE Transactions on*, vol. 57, no. 7, pp. 1417 –1430, july 2010.