ON QUANTIFYING THE EFFECTS OF NONCIRCULARITY ON THE COMPLEX FASTICA ALGORITHM

Mike Novey and Tülay Adalı

University of Maryland Baltimore County, Baltimore, MD 21250
{mnovey1, adali@umbc.edu}

ABSTRACT

The complex fast independent component analysis (c-FastICA) algorithm is one of the most popular methods for solving the ICA problem with complex-valued data. In this study, we extend the work of Bingham and Hyvärinen [1] by deriving conditions for local stability for the more general case of noncircular sources. We use the results of the analysis to quantify the effects of noncircularity on the performance of the algorithm using various nonlinearities and source distributions. Simulations are presented to demonstrate the results of our analysis.

Index Terms— Nonlinear estimation, stability

1. INTRODUCTION

Independent component analysis for separating complex-valued signals has found utility in many applications such as wireless communications [2] and radar [3], and data analysis in magnetic resonance imaging [4] and electroencephalograph [5]. Depending on the application, the sources may be both sub-Gaussian and super-Gaussian, and specifically in the complex domain, can have circular—rotation invariant—and noncircular—rotation variant distributions. Current work dealing with noncircular sources is given in [6, 7] where complex-valued kurtosis is used as the cost function. Another approach, the strong-uncorrelating transform (SUT) [8], uses second-order statistical information through the covariance and pseudo-covariance matrices and performs ICA by joint diagonalization of these matrices. Although efficient, the algorithm restricts the sources to be noncircular with distinct spectra of the pseudo-covariance matrix.

The complex FastICA (c-FastICA) algorithm of Bingham and Hyvärinen is one of the most popular methods for performing ICA when dealing with complex-valued sources. The algorithm does not restrict the cost to kurtosis, as in [6, 7], but uses general nonlinearities that are less susceptible to outliers. In [1], authors present a fixed-point algorithm and derive the conditions for local stability assuming circular sources.

In this study, we provide a rigorous local stability analysis of the c-FastICA algorithm without the circularity constraint, thereby allowing us to quantify the effects of noncircularity on stability. We find that the performance of c-FastICA is affected more severely for sub-Gaussian sources than super-Gaussian—which we show through analysis and simulations.

2. COMPLEX ICA

2.1. Complex preliminaries

A complex variable z is defined in terms of two real variables z^R and z^I as $z = z^R + jz^I$ with $\arg(z) = \text{atan}(z^I/z^R)$. Statistics of a complex random vector $\mathbf{x} = \mathbf{x}^R + j\mathbf{x}^I$ are defined by the joint probability density function (pdf) $p_{\mathbf{x}}(\mathbf{x}^R, \mathbf{x}^I)$. The expectation of a complex random vector $\mathbf{x}$ is then given with respect to this pdf and is written as $E\{\mathbf{x}\} = E\{\mathbf{x}^R\} + jE\{\mathbf{x}^I\}$.

The covariance matrix is written as $\text{cov}(\mathbf{x}) = E\{(\mathbf{x}-E\{\mathbf{x}\})(\mathbf{x}-E\{\mathbf{x}\})^H\}$ where H denotes conjugate transpose and the pseudo-covariance matrix is defined as $\text{pcov}(\mathbf{x}) = E\{(\mathbf{x} - E\{\mathbf{x}\})(\mathbf{x} - E\{\mathbf{x}\})^T\}$ where T denotes the transpose. These two quantities together define the second-order statistics of a complex random vector, and the random vector is second-order circular if $\text{pcov}(\mathbf{x}) = \mathbf{0}$. A stronger definition of circularity is based on the pdf of the complex random variable such that for any α, the pdf of x and $e^{j\alpha}x$ are the same [9]. The kurtosis of a zero mean complex random variable, often used as a quantitative measure of non-Gaussianity, is defined in [1] as $\text{kurt}(y) = E\{|y|^4\} - 2\left(E\{|y|^2\}\right)^2 - \left|E\{y^2\}\right|^2$ and reduces to

$$\text{kurt}_c(y) = E\{|y|^4\} - 2 \tag{1}$$

when y is circular with unit variance. We use the following definitions of complex random vectors

$$\begin{aligned} \mathbf{z} &= [z_1, z_2, \ldots, z_N]^T \in \mathbb{C}^N \\ \tilde{\mathbf{z}} &= [z_1, z_1^*, \ldots, z_N, z_N^*]^T \in \mathbb{C}^{2N} \end{aligned} \tag{2}$$

and $\mathbf{M} \in \mathbb{C}^{N\times N}$ and $\tilde{\mathbf{M}} \in \mathbb{C}^{2N\times 2N}$ for complex matrices throughout this work where $*$ denotes complex conjugation. These forms will be used when we define the complex gradient and Hessian of non-analytic functions as defined in [10].

2.2. ICA in the complex domain

In ICA, the observed data $\mathbf{z}$ are typically expressed as a linear combination of latent variables such that $\mathbf{z} = \mathbf{A}\mathbf{s}$ where $\mathbf{s} = [s_1, \ldots, s_N]^T$ is the column vector of latent sources, $\mathbf{z} = [z_1, \ldots, z_N]^T$ is the column vector of observed mixtures, and matrix $\mathbf{A}$ is the $N \times N$ mixing matrix assumed invertible. We assume that the sources and mixing matrix are complex valued. ICA then identifies the statistically independent sources given the observed mixtures typically by estimating a demixing matrix $\mathbf{W}$ so that the source estimates become $\mathbf{W}\mathbf{z}$. We assume without loss of generality that the sources have zero mean and unit variance, *i.e.*, $E\{\mathbf{s}\mathbf{s}^H\} = \mathbf{I}$. We do not, however, assume circular sources, hence $E\{\mathbf{s}\mathbf{s}^T\} \neq \mathbf{0}$. One can recover the original sources up to a complex scaling and permutation provided that at most one source is Gaussian.

A preliminary sphering or whitening of $\mathbf{z}$ is first performed through the transform $\mathbf{V}$, resulting in

$$\mathbf{x} = \mathbf{V}\mathbf{z} = \mathbf{V}\mathbf{A}\mathbf{s} \equiv \mathbf{B}\mathbf{s} \tag{3}$$

where $E\{\mathbf{x}\mathbf{x}^H\} = \mathbf{I}$ and $\mathbf{B} \equiv \mathbf{V}\mathbf{A}$. We find that the new mixing matrix, $\mathbf{B}$, is unitary due to $E\{\mathbf{x}\mathbf{x}^H\} = \mathbf{B}E\{\mathbf{s}\mathbf{s}^H\}\mathbf{B}^H = \mathbf{I}$.

Note however that the whitening step does not decorrelate the components, *i.e.,* the pseudo-covariance matrix is not diagonalized as is done in the SUT [8]. We estimate each source, y_k, separately by finding a vector $\mathbf{w}$ such that

$$y_k = \mathbf{w}_k^H \mathbf{x} = \mathbf{w}_k^H \mathbf{B}\mathbf{s} = \mathbf{q}_k^H \mathbf{s} \tag{4}$$

where $\mathbf{w}_k$ is a column of $\mathbf{W}$ and $\mathbf{q}_k = [0, \ldots, q_k, 0, \ldots]^T$ when $\mathbf{W}^H = \mathbf{B}$, *i.e.,* the optimal solution. Constraining the source estimates such that $E\{y_k y_k^*\} = 1$, also constrains the weights to $||\mathbf{w}||^2 = 1$, $\mathbf{q}_k = [0, \ldots, e^{j\theta}, 0, \ldots]^T$, and $\mathbf{W}$ to a unitary matrix due to the whitening transform.

3. STABILITY ANALYSIS

The c-FastICA cost function, defined in [1], is

$$J(\mathbf{w}) = E\left\{G\left(\left|\mathbf{w}^H\mathbf{x}\right|^2\right)\right\} \tag{5}$$

where $G : \mathbb{R} \rightarrow \mathbb{R}$ is a smooth even function and $\mathbf{w} \in \mathbb{C}^N$ with $||\mathbf{w}|| = 1$. The resulting optimization problem is formulated as

$$\mathbf{w}_{\text{opt}} = \arg \max_{||\mathbf{w}||^2=1} J(\mathbf{w}). \tag{6}$$

Several choices for G in (5) were proposed [1] and are: $G_1 = \sqrt{a_1 + y}$, $G_2 = \log(a_1 + y)$, and $G_3 = \frac{1}{2}y^2$, where a_1 is an arbitrary constant and is chosen as 0.1 in this work as in [1]. The functions G_1 and G_2 are slowly growing functions providing robust estimators, and G_3 is motivated by kurtosis given in (1). Note that the cost function given in (5) does not utilize the phase of the sources indicating that any noncircularity information is lost as is the case of c-FastICA.

3.1. Stability conditions

We now present the conditions required for the optimal solution to be a stable point, *i.e.,* the cost function (5) is a local maximum or minimum. We derive the conditions for local stability at the optimal solution similar to that of [1] except for two significant differences: 1) we do *not* assume circular sources and 2) we work in the complex domain using the definitions of the complex derivative of non-analytic functions found in [11, 12] and Taylor series expansion shown in [10].

The major result of this section, derived in Appendix A.1 and based on a second-order analysis, is that a local minimum (*resp.* maximum) is achieved when estimating source one given sources $(i = 2, 3, \ldots, N)$ when

$$\begin{aligned} E\{g(|s_1|^2) \quad &+ \quad |s_1|^2 g'(|s_1|^2) \, - \, |s_1|^2 g(|s_1|^2)\} \\ &\pm \quad |E\{s_i^2\}\beta| \, > 0 \qquad (\textit{resp.} \, < 0) \end{aligned} \tag{7}$$

where we have defined $\beta = E\{g'(|s_1|^2) s_1^{*2} e^{j2\theta}\}$ and used the notation $g(z) = dG(z)/dz$ and $g'(z) = dg(z)/dz$. Note that in (7), source one was chosen as an example to show the source to source stability dependence and must be true for all source combinations. For circular sources, the expression reduces to

$$E\left\{g(|s_1|^2) + |s_1|^2 g'(|s_1|^2) - |s_1|^2 g(|s_1|^2)\right\} > 0 \qquad (\textit{resp.} \, < 0) \tag{8}$$

which coincides with the result given in [1] since $E\{s_i^2\} = 0$. This result for circular sources shows that stability is guaranteed as long as one chooses to maximize or minimize the cost depending on the sign of equation (8). However for the noncircular case, the last term in (7) is not zero indicating that stability is not guaranteed since the inequality must be satisfied for both the plus and minus values.

3.2. Stability examples

Consider the case of applying the nonlinearity $G_3 = \frac{1}{2}y^2$, which is motivated by kurtosis (1), and assume circular sources. Substituting G_3 into (8) results in $2 - E\{|s_1|^4\} > 0$ or $\text{kurt}_c(s_1) < 0$, which implies that we minimize the cost for sub-Gaussian sources and maximize for super-Gaussian sources. Here, we define sub/super-Gaussianity with respect to the kurtosis of the sources.

Now let us examine stability when the sources are noncircular and again using the nonlinearity G_3. We first note that the term $|E\{s_i^2\}\beta|$ in (7) is not zero and expands to $|E\{s_i^2\}E\{s_1^{*2}\}|$, which, due to the unit variance constraint, $0 \leq |E\{s_i^2\}| \leq 1$. Equation (7) now becomes

$$\text{kurt}_c(s_1) \pm \left|E\{s_i^2\}E\{s_1^{*2}\}\right| < 0 \qquad (\textit{resp.} \, > 0) \tag{9}$$

implying that instability may result if we are minimizing or maximizing the cost based on the sign of kurtosis.

We illustrate the linear relationship shown in equation (9), for nonlinearity G_3, by plotting $|E\{s_i^2 s_1^{*2}\}|$ versus $\text{kurt}_c(s_1)$ in Figure 1. Figure 1 illustrates regions of instability as the sources become noncircular, *i.e.,* $|E\{s_i^2 s_1^{*2}\}| \rightarrow 1$. Note that for simplicity we assume $E\{s_i^2\} = 1$ and only vary $E\{s_1^{*2}\}$. What we glean from the figure is that the space of sub-Gaussian distributions for signals of engineering interest, $\text{kurt}_c(s_{\text{sub}}) \in [-1, 0)$, stability is always affected by noncircularity. We note that these signals of engineering interest include: quadrature amplitude modulation (QAM), binary phase shift keying (BPSK), functional magnetic resonance data, and complex sinusoids. However, super-Gaussian sources, $\text{kurt}_c(s_{\text{super}}) \in (0, \infty)$, are affected only if the kurtosis value is in the narrow range of zero to one.

The stability results for nonlinearities G_1 and G_2 are also shown in Figure 1. However since closed form solutions do not exist for the expectations in equation (7), numerical results were calculated using sources realized by the bivariate generalized Gaussian distribution (ggd) described in Appendix A.2. As seen in the figure, the regions of instability for G_1 and G_2 are marginally larger than that of G_3. We conclude from this section that the c-FastICA algorithm will provide good separation performance with noncircular sources specifically for super-Gaussian sources, however performance is limited for the sub-Gaussian case.

As a final example, we use complex BPSK sources that are intrinsically noncircular, $|E\{s^2\}| = 1$, and sub-Gaussian, $\text{kurt}_c(s) = -1$, with probability mass function (pmf)

$$p_{\text{bpsk}}(s) = 0.5\delta(s - \mu) + 0.5\delta(s + \mu) \tag{10}$$

where μ is one point in the two-point constellation, *i.e.,* $\mu = \cos(\theta) + j\sin(\theta)$, and is on the unit circle due to the unit variance constraint. Using (10), we are able to get a closed form solution to equation (7) for all three nonlinearities resulting in $1 \pm 1 > 0$. This result indicates that c-FastICA will be unstable when applied to BPSK sources—we do not show the derivation here due to space limitations.

4. SIMULATIONS

Our goal in this section is to demonstrate the results of our stability analysis, specifically those shown in Figure 1, by quantifying performance of c-FastICA with varying noncircularity. For our simulations we use the Amari index, $I_A \in [0, 1]$, defined in [13] as our performance measure where smaller values indicate better performance with zero indicating perfect separation. Eight sources are used with

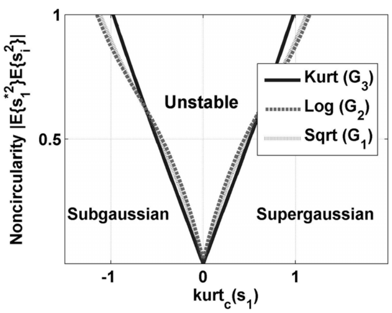
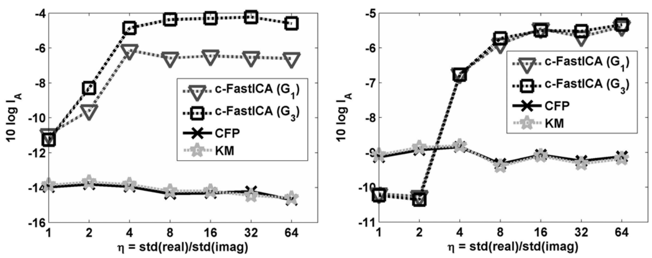


Fig. 1. Plot of $|E\{s_1^{*2}\}E\{s_i^2\}|$, noncircularity measure, versus kurtosis (1), with the areas of instability noted for three nonlinearities.



(a) Sub-Gaussian sources (b) Super-Gaussian

Fig. 2. Plot of separation performance, $10\log I_A$, versus noncircularity measure η.

results averaged over 100 runs. Our sources are realizations of the bivariate ggd distribution outlined in Appendix A.2. Sources are made noncircular by varying the real to imaginary asymmetry defined by $\eta = \sqrt{E\{(s^R)^2\}/E\{(s^I)^2\}}$, *i.e.,* the ratio of the standard deviations of the real and imaginary parts. We compare the results of c-FastICA with more recent algorithms that specifically address noncircularity and use kurtosis as the cost function. These are the complex fixed point (CFP) [6] and kurtosis maximization (KM) [7] algorithms.

Figure 2 highlights the results of our stability analysis by depicting the performance versus noncircularity measure η for c-FastICA using G_1 and G_3. Figure 2(a) displays the results with sub-Gaussian sources illustrating that as the sources become more noncircular, *i.e.,* $\eta \geq 4$, c-FastICA fails to separate as predicted by Figure 1. As expected, we see similar performance with KM and CFP due to both using kurtosis as the cost function and explicitly taking noncircularity into account. Figure 2(b) shows the results with super-Gaussian sources with very low kurtosis values, $\text{kurt}_c(s) \in [.25, 1)$. Again c-FastICA degrades as η increases. We also see how c-FastICA outperforms the kurtosis based algorithms KM and CFP when circular, *i.e.,* $\eta = 1$. This result indicates the advantage of being able to select a nonlinearity based on source statistics.

5. CONCLUSIONS

In this paper we provide a rigorous stability analysis of the c-FastICA algorithm to the more general case of noncircular sources. We use this analysis to quantify the effects of noncircularity on performance and find that c-FastICA's performance is degraded more severely with sub-Gaussian than with super-Gaussian sources. We show this result through analysis and simulations.

A. APPENDIX

A.1. Stability conditions of cost function (5)

We make the orthogonal change of coordinates $\mathbf{q} = \mathbf{A}^H\mathbf{w}$ resulting in the cost function $J(\mathbf{q}) = E\{G(yy^*)\}$ where $y = \mathbf{w}^H\mathbf{x} = \mathbf{q}^H\mathbf{s}$. Without loss of generality, we assume an optimal solution for s_1 at $\mathbf{q}_1 = [e^{j\theta}, 0, \dots]^T$ where θ points in the direction of the principal component of s_1 as shown in [14]. If s_1 is circular then θ is an arbitrary phase shift.

We seek a Taylor series expansion of J around the optimal solution $\mathbf{q}_1$, but unlike the approach in [1] using real-valued vectors, we choose to work in the less-cumbersome complex domain. The cost J is not analytic in $\mathbf{q}$ but is analytic in $\mathbf{q}$ and $\mathbf{q}^*$ independently. Because of this condition, we apply Wirtinger calculus [11, 12, 7] and the partial derivative can be found directly by differentiating with respect to $\mathbf{q}$ while treating $\mathbf{q}^*$ as a constant resulting in

$$\frac{\partial J}{\partial q_i} = g(yy^*)ys_i^*$$

where g is the derivative of G and we used the chain rule and $\partial q^*/\partial q = 0$. The second derivatives can be found similarly as

$$\begin{aligned}
a_{ik} &= E\left\{\frac{\partial^2 J}{\partial q_i^* \partial q_k}\right\} = E\{s_i s_k^* \left[g'(yy^*)yy^* + g(yy^*)\right]\} \\
b_{ik} &= E\left\{\frac{\partial^2 J}{\partial q_i^* \partial q_k^*}\right\} = E\{s_i s_k g'(yy^*)y^{*2}\} \\
c_{ik} &= E\left\{\frac{\partial^2 J}{\partial q_i \partial q_k}\right\} = b_{ik}^*
\end{aligned}$$

where g' is the derivative of g. We use the above derivative definitions and the complex gradient and Hessian defined in [10] to write the gradient as

$$\tilde{\nabla}_{\mathbf{q}} J = E\begin{pmatrix} \frac{\partial J}{\partial q_1} \\ \frac{\partial J}{\partial q_1^*} \\ \vdots \\ \frac{\partial J}{\partial q_N} \\ \frac{\partial J}{\partial q_N^*} \end{pmatrix} = \begin{pmatrix} E\{g(yy^*)s_1^*y\} \\ E\{g(yy^*)s_1y^*\} \\ \vdots \\ E\{g(yy^*)s_N^*y\} \\ E\{g(yy^*)s_Ny^*\} \end{pmatrix} \tag{11}$$

and Hessian

$$\begin{aligned}
\tilde{\mathbf{H}}_{\mathbf{q}} J &= E\left\{\frac{\partial^2 J}{\partial \tilde{\mathbf{q}}^* \partial \tilde{\mathbf{q}}^T}\right\} \\
&= \begin{pmatrix} a_{11} & b_{11} & a_{12} & b_{12} & \dots & a_{1N} & b_{1N} \\ b_{11}^* & a_{11} & b_{12}^* & a_{12} & \dots & b_{iN}^* & a_{1N} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ b_{N1}^* & a_{N1} & b_{N2}^* & a_{N2} & \dots & b_{NN}^* & a_{NN} \end{pmatrix}
\end{aligned} \tag{12}$$

where the vector format is defined in (2). Evaluating the gradient (11) and Hessian (12) at $\mathbf{q}_1$ and using the whiteness and independence of $\mathbf{s}$ we find

$$\tilde{\nabla}_{\mathbf{q}} J(\mathbf{q}_1) = \begin{pmatrix} E\{g(|s_1|^2)|s_1|^2\}e^{-j\theta} \\ E\{g(|s_1|^2)|s_1|^2\}e^{j\theta} \\ 0 \\ \vdots \\ 0 \end{pmatrix}$$

and

$$\tilde{\mathbf{H}}_{\mathbf{q}} J(\mathbf{q}_1) = \begin{pmatrix} a_{11} & b_{11} & \mathbf{0} & \dots & \mathbf{0} \\ b_{11}^* & a_{11} & \mathbf{0} & \dots & \mathbf{0} \\ 0 & 0 & \mathbf{B}_2 & \mathbf{0} & \vdots \\ \vdots & \vdots & \vdots & \ddots & \mathbf{0} \\ 0 & 0 & \mathbf{0} & \dots & \mathbf{B}_N \end{pmatrix} \quad (13)$$

where at the optimal solution $a_{11} = E\{|s_1|^4 g'(|s_1|^2) + |s_1|^2 g(|s_1|^2)\}$, $b_{11} = E\{|s_1|^4 g'(|s_1|^2) e^{j2\theta}\}$, and $\mathbf{B}_i$ is a 2×2 matrix defined as $\mathbf{B}_i = \begin{pmatrix} \gamma & E\{s_i^2\}\beta \\ E\{s_i^{*2}\}\beta^* & \gamma \end{pmatrix}$ where $\gamma = E\{g'(|s_1|^2)|s_1|^2 + g(|s_1|^2)\}$ and $\beta = E\{g'(|s_1|^2) e^{j2\theta} s_1^{*2}\}$.

We now make a small perturbation, $\boldsymbol{\varepsilon} = [\varepsilon_1, \varepsilon_2, \dots, \varepsilon_N]$, around the optimal solution $\mathbf{q}_1$ using the complex Taylor series expansion derived in [10] as

$$\begin{aligned} J(\mathbf{q}_1 + \boldsymbol{\varepsilon}) &= J(\mathbf{q}_1) + \tilde{\boldsymbol{\varepsilon}}^T \tilde{\nabla}_{\mathbf{q}} J(\mathbf{q}_1) + \frac{1}{2} \tilde{\boldsymbol{\varepsilon}}^H \tilde{\mathbf{H}}_{\mathbf{q}} J(\mathbf{q}_1) \tilde{\boldsymbol{\varepsilon}} + o(||\boldsymbol{\varepsilon}||^2) \\ &= J(\mathbf{q}_1) + E\{g(|s_1|^2)|s_1|^2\}(\varepsilon_1 e^{-j\theta} + \varepsilon_1^* e^{j\theta}) \; + \\ & \quad |\varepsilon_1|^2 a_{11} + \frac{1}{2}(\varepsilon_1^2 b_{11}^* + \varepsilon_1^{*2} b_{11}) + \gamma \sum_{i>1}^N |\varepsilon_i|^2 + \\ & \quad \frac{1}{2}\beta \sum_{i>1}^N \varepsilon_i^2 E\{s_i^2\} + \frac{1}{2}\beta^* \sum_{i>1}^N \varepsilon_i^2 E\{s_i^{*2}\} + o(||\boldsymbol{\varepsilon}||^2). \end{aligned}$$

Noting that $||\mathbf{q}_1 + \boldsymbol{\varepsilon}||^2 = 1 + e^{-j\theta}\varepsilon_1 + e^{j\theta}\varepsilon_1^* + \sum_{i=1}^N |\varepsilon_i|^2$ and the constraint, $||\mathbf{q}|| = 1$, we obtain

$$e^{-j\theta}\varepsilon_1 + e^{j\theta}\varepsilon_1^* = -\sum_{i=1}^N |\varepsilon_i|^2. \quad (14)$$

Substituting (14) into the Taylor series expansion we get

$$\begin{aligned} J(\mathbf{q}_1 + \boldsymbol{\varepsilon}) = J(\mathbf{q}_1) + |\varepsilon_1|^2 E\{|s_1|^4 g'(|s_1|^2)\} + \frac{1}{2}\varepsilon_1^2 b_{11}^* + \frac{1}{2}\varepsilon_1^{*2} b_{11} + \\ E\{|s_1|^2 g'(|s_1|^2) + g(|s_1|^2) - g(|s_1|^2)|s_1|^2\} \sum_{i>1}^N |\varepsilon_i|^2 + \\ \frac{1}{2}\sum_{i>1}^N \left(\varepsilon_i^2 E\{s_i^2\}\beta^* + \varepsilon_i^{*2} E\{s_i^{*2}\}\beta\right) + o(||\boldsymbol{\varepsilon}||^2) \end{aligned}$$

where the term $|\varepsilon_1|^2$ is of order $o(||\boldsymbol{\varepsilon}||^2)$ according to (14) and can be neglected, resulting in

$$\begin{aligned} J(\mathbf{q}_1 + \boldsymbol{\varepsilon}) &= J(\mathbf{q}_1) + E\{|s_1|^2 g'(|s_1|^2) + g(|s_1|^2) - \\ & \quad g(|s_1|^2)|s_1|^2\} \sum_{i>1}^N |\varepsilon_i|^2 + \\ & \quad \sum_{i>1}^N |\varepsilon_i|^2 |E\{s_i^2\}|\beta| \cos(\varphi_i) + o(||\boldsymbol{\varepsilon}||^2) \quad (15) \end{aligned}$$

where $\varphi_i = \arg(\varepsilon_i^2) + \arg(E\{s_i^2\}) + \arg(\beta)$ is some arbitrary phase shift and we used the identity $z + z^* = 2z^R = 2|z|\cos[\arg(z)]$ in the last line. The term $\cos(\varphi_i) \in [-1, 1]$ for any perturbation ε_i, resulting in the conditions for a local minimum (*resp.* maximum) as

$$\begin{aligned} E\{g(|s_1|^2) \quad &+ \quad |s_1|^2 g'(|s_1|^2) \; - \; |s_1|^2 g(|s_1|^2)\} \\ &\pm \quad |E\{s_i^2\}\beta| \; > 0 \qquad (\textit{resp.} \; < 0) \end{aligned}$$

which must be satisfied for each source s_i, where $i = [2, 3, \dots, N]$. The result outlined in (15) is identical to that in [1] for circular sources, *i.e.,* the term $E\{s_i^2\} = 0$ when source s_i is circular.

A.2. Generalized Gaussian distribution

We modify the bivariate ggd model from [15] for the noncircular unit variance case as

$$p_{\text{ggd}}(s^R, s^I, p, m) = \frac{\Gamma(4/p)p}{2\pi\Gamma(2/p)^2\sqrt{1-m^2}} e^{-\left[\frac{(s^R)^2}{m+1} + \frac{(s^I)^2}{1-m}\right]^{p/2} \gamma^{p/2}}$$

where p is the shape parameter, $\gamma = \left[\frac{\Gamma(4/p)}{\Gamma(2/p)}\right]$, $m = E\{s^2\}$ and $E\{ss^*\} = 1$. By adjusting the shape parameter, we can generate Gaussian variates with $p = 2$, sub-Gaussian with $p > 2$, and super-Gaussian with $0 < p < 2$. To generate these complex random variables in Matlab, we modify the approach in [16] for real-valued variates to $s = \text{gamrnd}(2/p, 1)^{1/p} \exp(j2\pi \; \text{rand})$ where $\text{gamrnd}(2/p, b)$ generates gamma random variables with shape parameter $2/p$ and rand generates uniformly distributed variables $\in [0, 1]$. The sources are then divided by their standard deviations to yield unit variance.

B. REFERENCES

[1] E. Bingham and A. Hyvärinen, "A fast fixed-point algorithm for independent component analysis of complex valued signals," *Int. J. Neural Systems*, vol. 10, pp. 1–8, 2000.

[2] K. Waheed and F. Salem, "Blind information-theoretic multiuser detection algorithms for DS-CDMA and WCDMA downlink systems," *IEEE Trans. on Neural Networks*, vol. 16, no. 4, pp. 937–948, July 2005.

[3] J.-F. Cardoso and A. Souloumiac, "Blind beamforming for non-gaussian signals," in *IEE Proc. Radar Signal Proc.*, 1993, vol. 140, pp. 362–370.

[4] V. D. Calhoun and T. Adalı, "Unmixing fMRI with independent component analysis," *IEEE Engineering in Medicine and Biology Magazine*, vol. 25, pp. 79–90, April 2006.

[5] J. Annemüller, T. J. Sejnowski, and S. Makeig, "Complex spectral domain independent component analysis of electrocephalographic data," in *Proc.* ICA *Workshop*, Nara, Japan, March 2003.

[6] Scott C. Douglas, "Fixed-point algorithms for the blind separation of arbitrary complex-valued non-gaussian signal mixtures," *EURASIP J. Appl. Signal Process.*, vol. 2007, no. 1, pp. 83–83, 2007.

[7] H. Li and T. Adalı, "Gradient and fixed-point complex ICA algorithms based on kurtosis maximization," in *Proc. IEEE Workshop on Machine Learning for Signal Processing (MLSP)*, Maynooth, Ireland, Sept. 2006, pp. 85–90.

[8] J. Eriksson and V. Koivunen, "Complex-valued ICA using second-order statistics," in *Proc. IEEE Workshop on Machine Learning for Signal Processing (MLSP)*, Sao Luis, Brazil, 2004, pp. 183–192.

[9] B. Picinbono, "On circularity," *IEEE Trans. Signal Processing*, vol. 42, pp. 3473–3482, Dec. 1994.

[10] A. van den Bos, "Complex gradient and Hessian," in *IEE Proc. Image Signal Processing*, Dec. 1994, vol. 141, pp. 380–382.

[11] W. Wirtinger, "Zur formalen theorie der funktionen von mehr complexen veränderlichen," *Math. Ann.*, vol. 97, pp. 357–375, 1927.

[12] D. H. Brandwood, "A complex gradient operator and its application in adaptive array theory," *Proc. Inst. Elect. Eng.*, vol. 130, no. 1, pp. 11–16, Feb. 1983.

[13] S. Amari, A. Cichocki, and H. H. Yang, "A new learning algorithm for blind signal separation," in *Advances in Neural Information Processing Systems*, D. S. Touretzky, M. C. Mozer, and M. E. Hasselmo, Eds. 1996, vol. 8, pp. 757–763, The MIT Press.

[14] M. Novey and T. Adalı, "ICA by maximization of nongaussianity using complex functions," in *Proc. IEEE Workshop on Machine Learning for Signal Processing (MLSP)*, Mystic, CT, 2005.

[15] M. Coban and R. Mersereau, "Adaptive subband video coding using bivariate generalized gaussian distribution model," in *Proc. ICASSP*, Atlanta, GA, May 1996, vol. 4, pp. 1990–1993.

[16] M. E. Johnson, "Computer generation of the exponential power distributions," in *J. Statistical Computation and Sim.*, 1979, vol. 9, pp. 239–240.