

Contrast Functions for Non-circular and Circular Sources Separation in Complex-Valued ICA

Zhe Chen and Jinwen Ma

Abstract—In this paper, the complex-valued ICA problem is studied in the context of blind complex-source separation. We formulate the complex ICA problem in a general setting, and define the superadditive functional that may be used for constructing a contrast function for circular complex sources separation. We propose several contrast functions and study their properties. Finally, we also discuss relevant issues and present the convex analysis of a specific contrast function.

I. INTRODUCTION

Independent component analysis (ICA) is a powerful tool in statistical signal and image processing [1]-[4], and it has also been an intensive research topic in neural network community. However, most of ICA research thus far focused on real domain. Recently, many research efforts attempted to generalize the ICA concept to complex domain, e.g. [5]-[14]. Because there are numerous ICA algorithms and related theories, it is imagined that there are also several routes that lead to complex ICA. Despite their differences, we category two important approaches here

- Complex ICA based on second-order statistics (SOS) or *generalized eigenvalue decomposition* (GEVD): In this approach, generalization from real to complex domain is relatively straightforward by replacing symmetric covariance matrix with Hermitian covariance matrix. Examples [2, 5] include AMUSE and SOBI (SOS-based), and FOBI, and JADE (GEVD-based). Specifically, SOS was found useful for separating complex circular Gaussian sources which have different spectra [10, 11].
- Complex ICA based on high-order statistics (HOS): In this approach, nonlinearity is used to produce high-order decorrelation. Example of such a class include adaptive algorithms such as complex FastICA [6] and complex Infomax [7, 8, 13].

Because of the unique nature of complex random variables, complex analysis has many inherent properties (such as differentiability, nonlinearity, etc.) that are different from real analysis. It is observed that most available research results are rather preliminary, a deep theoretical understanding of complex ICA, is far from complete. The current contribution may be viewed as an effort along this research line.

In this paper, we propose and derive suitable contrast functions for the complex ICA problem in the simultaneous

The authors are with the Laboratory for Advanced Brain Signal Processing (Z.C.) and Laboratory for Mathematical Neuroscience (J.M.), RIKEN Brain Science Institute, Wako-shi, Saitama 351-0198, Japan (email: {zhechen,jwma}@brain.riken.jp). J.M. is also with the Department of Information Science, School of Mathematical Sciences and LMAM, Peking University, Beijing 100871, China.

separation setup. Non-circular and circular complex sources are both discussed, but attention is focused on the circular sources. To sidestep the difficulty of constructing generic discriminating contrast functions in complex domain, we construct superadditive functions of class II [19] for circular complex variables, and show how to extend to complex domain as discriminating contrast functions. Our paper is rather theoretically oriented and currently contains no experimental study; we also present a convex analysis on the contrast function for a two-by-two source separation problem.

The rest of the paper is organized as follows. Section II presents mathematical preliminaries and an overview of important concepts. Section III describes the problem of complex ICA in a general setting. In Section IV, we define the criteria for contrast functions in complex ICA, and propose several contrast functions for circular complex sources. Section V presents a case study analysis of local or global minima/maxima of a specific contrast function. This is further followed by discussion in Section VI and concluding remarks in Section VII.

II. MATHEMATICAL PRELIMINARIES

A complex random variable $z \in \mathbb{C}$ is defined as $z = z_{\text{Re}} + jz_{\text{Im}}$, where $j = \sqrt{-1}$, and the real part $z_{\text{Re}} \in \mathbb{R}$ and imaginary part $z_{\text{Im}} \in \mathbb{R}$ are both real-valued random variables. By “complex-valued” here we mean the random number is strictly complex if not stated otherwise; namely, the variable’s imaginary part is not almost sure zero. For a complex-valued $z = z_{\text{Re}} + jz_{\text{Im}}$, its complex conjugate is defined by $z^* = z_{\text{Re}} - jz_{\text{Im}}$. The relationship $z = z^*$ holds if and only if $z_{\text{Im}} = 0$. The complex-valued variable can also be represented as $z = |z|e^{j\theta}$, with *modulus* $|z| = \sqrt{z_{\text{Re}}^2 + z_{\text{Im}}^2}$ and the *phase* $\theta = \arg(z)$ ($0 \leq \theta < 2\pi$). The statistical properties of $z \in \mathbb{C}$ are characterized by the *joint* probability density function (pdf) of z_{Re} and z_{Im} : $p(z) = p(z_{\text{Re}}, z_{\text{Im}}) \in \mathbb{R}$, provided that it exist. Let $\tilde{\mathbf{z}} = [z_{\text{Re}}, z_{\text{Im}}]^T$, then $p(z)$ can be written as a form of $p(\tilde{\mathbf{z}})$ with its associated first and second cumulant statistics defined as $\mathbb{E}[\tilde{\mathbf{z}}]$ and $\Sigma_z = \text{cov}[\tilde{\mathbf{z}}]$

$$\Sigma_z = \begin{bmatrix} \mathbb{E}[z_{\text{Re}}^2] - \mathbb{E}^2[z_{\text{Re}}] & \mathbb{E}[z_{\text{Re}}z_{\text{Im}}] - \mathbb{E}[z_{\text{Re}}]\mathbb{E}[z_{\text{Im}}] \\ \mathbb{E}[z_{\text{Re}}z_{\text{Im}}] - \mathbb{E}[z_{\text{Re}}]\mathbb{E}[z_{\text{Im}}] & \mathbb{E}[z_{\text{Im}}^2] - \mathbb{E}^2[z_{\text{Im}}] \end{bmatrix}.$$

Consequently, the Shanon (differential) entropy of the complex random variable z , denoted as $H(z)$, is defined as the *joint* entropy of the real and imaginary components:

$$H(z) = H(z_{\text{Re}}, z_{\text{Im}}) = H(z_{\text{Re}}|z_{\text{Im}}) + H(z_{\text{Im}}).$$

Notably, the entropy $H(z)$ is a quantity that is independent on the mean value of z .

Given an appropriate probability metric of the random complex variable z , we can identify and calculate the first and second-order moment statistics:

- The first-order moment (mean):

$$\mathbb{E}[z] = \mathbb{E}[z_{\text{Re}} + jz_{\text{Im}}] = \mathbb{E}[z_{\text{Re}}] + j\mathbb{E}[z_{\text{Im}}].$$

- The second-order moment:

$$\mathbb{E}[z^2] = \mathbb{E}[z_{\text{Re}}^2] - \mathbb{E}[z_{\text{Im}}^2] + 2j\mathbb{E}[z_{\text{Re}}z_{\text{Im}}].$$

- The second-order cumulant (variance):

$$\text{var}[z] = \mathbb{E}[|z - \mathbb{E}[z]|^2] = \mathbb{E}[|z|^2] - |\mathbb{E}[z]|^2.$$

The covariance of two complex random variables z_i and z_j is defined as $C_{ij} = \mathbb{E}[(z_i - \mathbb{E}[z_i])(z_j^* - \mathbb{E}[z_j^*])] = \mathbb{E}[z_i z_j^*] - \mathbb{E}[z_i]\mathbb{E}[z_j^*]$. Two complex random variables z_i and z_j ($j \neq i$) are said to be mutually *uncorrelated* if $C_{ij} = 0$.

The above definitions can be generalized to complex vectors. Let $\mathbf{z} = [z_1, \dots, z_n]$ be a complex-valued random vector, and let $\mathbf{z}^H = [z_1^*, \dots, z_n^*]^T \equiv (\mathbf{z}^*)^T$ be its Hermitian transpose (i.e., conjugate plus transpose), then we can define its mean $\mathbb{E}[\mathbf{z}]$, and covariance matrix $\text{cov}[\mathbf{z}] \equiv \mathbb{E}[(\mathbf{z} - \mathbb{E}[\mathbf{z}])(\mathbf{z} - \mathbb{E}[\mathbf{z}])^H]$; in addition, the *pseudo-covariance matrix* can be defined as $\text{pcov}[\mathbf{z}] \equiv \mathbb{E}[(\mathbf{z} - \mathbb{E}[\mathbf{z}])(\mathbf{z} - \mathbb{E}[\mathbf{z}])^T]$.

Definition 1: A complex random variable z is defined as “circular” if for any real-valued number α , the pdfs of $p(z)$ and $p(e^{j\alpha}z)$ are the same (i.e., $p(z)$ is rotation invariant).

The complex-valued random vector \mathbf{z} is called *second-order circular* if its pseudo-covariance matrix is a null matrix (i.e., with all entries as zeros); if $\mathbb{E}[z] = 0$, then the second-order circularity implies that $\mathbb{E}[z^2] = 0$ and real and imaginary parts of z are uncorrelated and have equal variances. If $\mathbb{E}[\mathbf{z}\mathbf{z}^H]$ is diagonal, then we say \mathbf{z} is uncorrelated; \mathbf{z} is called *strongly uncorrelated* if $\mathbb{E}[\mathbf{z}\mathbf{z}^H]$ and $\mathbb{E}[\mathbf{z}\mathbf{z}^T]$ are both diagonal. When the real and imaginary parts of \mathbf{z} have equal variance, \mathbf{z} is often said to be *symmetric*; if \mathbf{z} is symmetric and strongly uncorrelated, then \mathbf{z} is second-order circular, namely $\mathbb{E}[\mathbf{z}\mathbf{z}^T] = \mathbf{0}$.

Similarly, the skewness and kurtosis statistics for a *zero-mean* complex random variable z can be defined [1]:

$$\text{skewness}(z) = \mathbb{E}[|z|^3] / (\mathbb{E}[|z|^2])^{3/2}, \quad (1)$$

$$\text{kurtosis}(z) = \mathbb{E}[|z|^4] - 2(\mathbb{E}[|z|^2])^2 - |\mathbb{E}[z^2]|^2. \quad (2)$$

Definition 2: Two complex random variables z_1 and z_2 are said to be mutually independent if $p(z_1, z_2) = p(z_1)p(z_2)$; if z_1 and z_2 are both circular, then they are mutually independent if $p(|z_1|, |z_2|) = p(|z_1|)p(|z_2|)$.

Consider an optimization problem in the complex domain. Let $J(\mathbf{z})$ denote a real-valued, bounded scalar function with an argument of a complex-valued vector $\mathbf{z} \in \mathbb{C}^N$, the stationary point is described by equating the gradient operator to zero:

$$\nabla J \equiv \frac{\partial J(\mathbf{z})}{\partial \mathbf{z}^*} = \frac{1}{2} \left(\frac{\partial J(\mathbf{z})}{\partial z_{\text{Re}}} + j \frac{\partial J(\mathbf{z})}{\partial z_{\text{Im}}} \right) = 0.$$

which implies that at stationary points, $\frac{\partial J(\mathbf{z})}{\partial x_{\text{Re}}} = \frac{\partial J(\mathbf{z})}{\partial x_{\text{Im}}} = 0$.

Definition 3: A real-valued function $J(\mathbf{z})$ (where $\mathbf{z} \in \mathbb{C}^N$) is said to be convex in the complex plane if

$$J(\lambda \mathbf{z}_1 + (1 - \lambda)\mathbf{z}_2) \leq \lambda J(\mathbf{z}_1) + (1 - \lambda)J(\mathbf{z}_2)$$

for all $\mathbf{z}_1, \mathbf{z}_2 \in \mathbb{C}^N$ and $0 \leq \lambda \leq 1$.

Alternatively, if the Hessian matrix, $\mathbf{H} = \frac{\partial^2 J(\mathbf{z})}{\partial \mathbf{z} \partial \mathbf{z}^H}$, is positive semidefinite (i.e., with nonnegative real eigenvalues), then $J(\mathbf{z})$ is a convex function.

III. COMPLEX-VALUED ICA

In a similar vein in real-valued ICA, let us further consider a complex version of standard ICA model: $\mathbf{x} = \mathbf{A}\mathbf{s}$, where $\mathbf{s} \in \mathbb{C}^n$ denotes the n -dimensional complex-valued, elementwise-independent source vector, $\mathbf{x} \in \mathbb{C}^n$ denotes the n -dimensional complex-valued vector of mixture signals, and $\mathbf{A} \in \mathbb{C}^{n \times n}$ denotes a nonsingular (i.e., with full rank) complex-valued mixing matrix. It is noted that there are three types of indeterminacies arisen in complex-valued ICA:

- Permutation indeterminacy;
- Sign and scaling indeterminacy;
- Phase indeterminacy.

The first two indeterminacies are shared with the real-valued ICA; whereas the phase ambiguity arises from the inherent nature of the complex-valued data. The second and third indeterminacies, when combined together, is referred to complex scale ambiguity.

In the context of BSS, there are two kinds of approaches: (i) sequential extraction (deflation) approach, which extracts the independent components one by one in order (thereby referred to as “one-unit ICA”; and (ii) simultaneous separation approach, which separates all independent sources at the same time. In this paper, we will focus on the simultaneous separation approach; the one-unit complex ICA can be regarded as a special case of simultaneous separation as in real domain [20].

In terms of simultaneous separation, the goal of complex ICA is to find an unmixing matrix \mathbf{W} , to recover the original source signal vector \mathbf{s} , up to the above-mentioned three indeterminacies. Specifically, the separated output $\mathbf{y} \in \mathbb{C}^{n \times 1}$ is given by

$$\mathbf{y} = \mathbf{W}\mathbf{x} = (\mathbf{W}\mathbf{A})\mathbf{s} \equiv \mathbf{R}\mathbf{s}, \quad (3)$$

where we have defined $\mathbf{R} = \mathbf{W}\mathbf{A}$ for the later use. If we assume $\mathbb{E}[\mathbf{s}] = \mathbf{0}$ and $\mathbb{E}[\mathbf{s}\mathbf{s}^H] = \mathbf{I}$, it follows that $\mathbb{E}[\mathbf{y}] = \mathbf{R}\mathbb{E}[\mathbf{s}] = \mathbf{0}$, and $\mathbb{E}[\mathbf{y}\mathbf{y}^H] = \mathbf{R}\mathbf{R}^H$. In order to achieve source separation in a blind fashion, similar to the real-valued case, we minimize a contrast function such that in the global (or local) minima, the desirable solution $\mathbf{W} = \mathbf{A}^{-1}$ is obtained (or up to scale, permutation, and phase ambiguities). In this sense, we may write the separated signals as $\mathbf{y} = \mathbf{P}\mathbf{D}\mathbf{s}$, where \mathbf{D} is a complex-valued diagonal matrix: $\mathbf{D} = \text{diag}\{\alpha_1 e^{j\beta_1}, \dots, \alpha_n e^{j\beta_n}\}$ (where $\alpha_i, \beta_i \in \mathbb{R}$), and \mathbf{P} denotes the permutation matrix. The design of appealing contrast function is an essential topic in the ICA or BSS community (e.g., [4, 15, 20]); however, most researches

thus far are limited in real domain. We will extend the discussion to the context of complex ICA in the next section.

Without loss of generality and for discussion simplicity, we assume $\mathbb{E}[\mathbf{x}] = 0$ and $\mathbb{E}[\mathbf{x}\mathbf{x}^H] = \mathbf{I}$, otherwise the condition can be fulfilled by a preprocessing procedure known as the *strong-uncorrelating transform* [10, 11].

Before discussing specific contrast functions, let us first consider a generic contrast function known as *mutual information*, which is equivalent to the popular Kullback-Leibler divergence criterion that measures the discrepancy between the joint probability and the product of marginal probabilities:

$$\begin{aligned} I(y_1, \dots, y_n) &= \mathbb{E}_{p(\mathbf{y})} \left[\log \frac{p(\mathbf{y})}{\prod_{i=1}^n p(y_i)} \right] \\ &= \int p(\mathbf{y}) \log \frac{p(\mathbf{y})}{\prod_{i=1}^n p(y_i)} d\mathbf{y} \quad (4) \end{aligned}$$

Equation (4) is nonnegative and equals to zero if and only if the random variables $\{y_1, \dots, y_n\}$ are mutually independent. Notably if $\{y_1, \dots, y_n\}$ are jointly complex Gaussian, then (4) can also be rewritten as

$$I(y_1, \dots, y_n) = -\frac{1}{2} \log \left(\frac{\det(\mathbf{C}_y)}{\prod_{i=1}^n C_{ii}} \right) = -\frac{1}{2} \sum_{i=1}^n \log(\lambda_i) \quad (5)$$

where $C_{ij} = \mathbb{E}[(y_i - \mathbb{E}[y_i])(y_j^* - \mathbb{E}[y_j^*])]$, and λ_i are the eigenvalues obtained from a GEVD procedure: $\mathbf{C}_y \mathbf{u} = \lambda \mathbf{\Lambda} \mathbf{u}$, where $\mathbf{C}_y = \text{cov}[\mathbf{y}]$, and $\mathbf{\Lambda} = \text{diag}\{C_{11}, C_{22}, \dots, C_{nn}\}$.

Recall that

$$I(y_1, \dots, y_n) = \sum_{i=1}^n H(y_i) - H(\mathbf{y}),$$

where $H(\mathbf{y})$ denotes the joint entropy of the complex-valued random variables $\{y_1, \dots, y_n\}$; because $H(\mathbf{y}) = H(\mathbf{x}) + \log |\det(\mathbf{W})|$, equation (4) can be rewritten as

$$J(\mathbf{W}) = \sum_{i=1}^n H(y_i) - \log |\det(\mathbf{W})| - H(\mathbf{x}), \quad (6)$$

where $H(\mathbf{x})$ is a term that is irrelevant to the optimization with respect to (w.r.t.) \mathbf{W} . Taking the derivative of (6) w.r.t. \mathbf{W}^* yields the gradient operator

$$\nabla_{\mathbf{W}^*} J(\mathbf{W}) = \left(\mathbb{E}_{\mathbf{y}} [\psi(\mathbf{y}) \mathbf{y}^H] - \mathbf{I} \right) \mathbf{W}^{-H}, \quad (7)$$

where \mathbf{W}^{-H} denotes the Hermitian transpose of \mathbf{W}^{-1} , and $\psi(\mathbf{y}) = [\psi(y_1), \dots, \psi(y_n)]^T$ is a complex vector function defined by an elementwise operation:

$$\begin{aligned} \psi(y_i) &= -\frac{d \log p(y_i)}{dy_i^*} = -\frac{\frac{\partial p(y_i)}{\partial y_i^{\text{Re}}} + j \frac{\partial p(y_i)}{\partial y_i^{\text{Im}}}}{p(y_i)} \\ &= \frac{\partial \log p(y_i)}{\partial y_i^{\text{Re}}} + j \frac{\partial \log p(y_i)}{\partial y_i^{\text{Im}}}, \quad (8) \end{aligned}$$

where $\psi(\cdot)$ is known as the *complex score function*. However, $p(y_i) \equiv p(y_i^{\text{Re}}, y_i^{\text{Im}})$ invokes a joint probability density function that complicates the calculation of (8) because y_i^{Re} and y_i^{Im} are generally not mutually independent.

From (6), applying the ‘‘natural gradient’’ trick (see [2]) would yield the stochastic complex-valued natural gradient rule [7, 13]:

$$\Delta \mathbf{W} = \eta (\mathbf{I} - \psi(\mathbf{y}) \mathbf{y}^H) \mathbf{W}. \quad (9)$$

In light of the *Liouville’s theorem*, we know that there is a tradeoff between the *boundedness* and *analyticity* in the choice of nonlinearity in complex domain; many research efforts have been devoted to finding a proper $\psi(\cdot)$ (e.g., [7]-[9], [13], [14]); however, the reported methods are rather ad hoc or heuristic and rigorous theoretical justification still remains unclear.

To sidestep such a difficulty, we may partially convert the complex-valued problem into real domain. Specifically, we may introduce an auxiliary complex random variable as $e_i = y_i^{\text{Re}} - y_i^{\text{Im}}$, and further define the conditional entropy of $H(y_i^{\text{Re}} | y_i^{\text{Im}})$ as

$$H(y_i^{\text{Re}} | y_i^{\text{Im}}) = H(y_i^{\text{Re}} - y_i^{\text{Im}}) \equiv H(e_i) \quad (10)$$

Therefore, the entropy $H(y_i) \equiv H(y_i^{\text{Re}}, y_i^{\text{Im}})$ can be represented by

$$\begin{aligned} H(y_i) &= H(y_i^{\text{Im}} | y_i^{\text{Re}}) + H(y_i^{\text{Re}}) \equiv H(-e_i) + H(y_i^{\text{Re}}) \\ &= H(y_i^{\text{Re}} | y_i^{\text{Im}}) + H(y_i^{\text{Im}}) \equiv H(e_i) + H(y_i^{\text{Im}}) \quad (11) \end{aligned}$$

Notably e_i is generally non-Gaussian; if y_i^{Re} and y_i^{Im} are both Gaussian and with zero mean and variance 1/2, then e_i will also be Gaussian, with zero mean and variance 1. In light of (10), equation (6) may be rewritten as

$$\begin{aligned} J(\mathbf{W}) &= \sum_{i=1}^n H(y_i) - \log |\det(\mathbf{W})| \\ &= \sum_{i=1}^n H(e_i) + \sum_{i=1}^n H(y_i^{\text{Im}}) - \log |\det(\mathbf{W})| \quad (12) \end{aligned}$$

By treating e_i and y_i^{Im} as two separate real-valued random variables, we may approximate their differential entropy using various approximate estimators (e.g., [4, 14, 15]). A brief derivation and description of approximate estimate of $\frac{\partial J}{\partial w_{i,j}^*}$ of (12), based on Gram-Charlier expansion, is given in the Appendix.

IV. CONTRAST FUNCTIONS FOR CIRCULAR SOURCES

In this section, we focus on a restricted class of complex-valued ICA for *circular* complex random variables, whose pdf is merely related to the modulus (i.e., phase independent). Note that for zero-mean circular signals, the pdf $p(z) = g(|z|)$ is *radially isotropic*; hence, although $p(z)$ is defined as a two-dimensional pdf in complex plane, it has only one degree of freedom (i.e., $|z|$), and we only need to estimate its one-dimensional profile (i.e., the centrifugal direction along the origin). Figure 1 illustrates one example of circular complex Gaussian pdf and another example of circular complex sub-Gaussian pdf; their one-dimensional centrifugal profiles correspond to the one-dimensional Gaussian and two-modal Gaussian mixture, respectively.

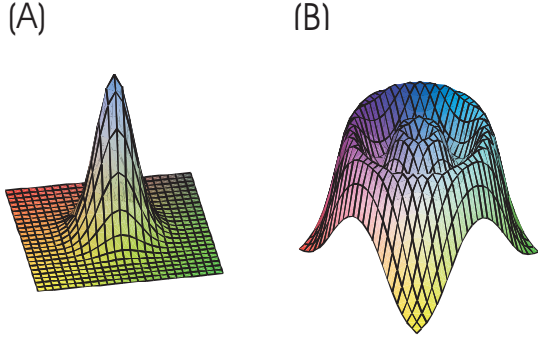


Fig. 1. The graphical plots of $p(|z|)$ in the complex plane $|z| = \sqrt{z_{\text{Re}}^2 + z_{\text{Im}}^2}$. (A) Isotropic pdf Gaussian circular signal. (B) Isotropic pdf sub-Gaussian (two-modal) circular signal.

Let us assume all the sources are circular and their distributions have a form of $p_s(s_i) = g(|s_i|)$, where $g(\cdot)$ denotes a specific nonlinear function associated with the i th complex-valued source s_i ; and the associated score function has the form $\psi(s) = -\frac{s}{|s|} \frac{g'(|s|)}{g(|s|)}$. In particular, given an arbitrary nonzero complex number $\alpha \in \mathbb{C}$ and a circular complex random variable s , we have

$$p_{\alpha s}(\alpha s) = \frac{1}{|\alpha|} p_s(s) = \frac{1}{|\alpha|} g(|s|). \quad (13)$$

and correspondingly,

$$\begin{aligned} H(\alpha s) &= -\int p_{\alpha s}(\alpha s) \log p_{\alpha s}(\alpha s) d|\alpha s| \\ &= -\frac{1}{|\alpha|} \int p_s(s) (\log p_s(s) - \log |\alpha|) d|\alpha s| \\ &= H(s) + \log |\alpha| \end{aligned} \quad (14)$$

$$\exp(H(\alpha s)) = |\alpha| \exp(H(s)). \quad (15)$$

In complex ICA, a contrast function is a *functional* of the distribution of \mathbf{y} , which is minimized when $\mathbf{R} = \mathbf{W}\mathbf{A}$ equals to the product of a permutation and a complex diagonal matrix. Following Comon [3] and Pham [15], we call a contrast function *discriminating* if it attains its minimum only when the successful separation is achieved. The essential goal is then to design discriminant contrast functions that lead to feasible solutions (either local or global minima) of complex ICA. Here, we restrict our attention on the contrast function for separating independent circular complex signals.

To characterize the identifiability of the complex ICA model, the complex analogue of the well-known *Darmonis-Skitovich Theorem*, which is fundamental to the concept of ICA [3], is stated here [10]:

Theorem 1: Let s_1, \dots, s_n be mutually independent complex random variables. For $\alpha_i, \beta_i \in \mathbb{C}$ ($i = 1, \dots, n$), if the linear forms $x_1 = \sum_{i=1}^n \alpha_i s_i$ and $x_2 = \sum_{i=1}^n \beta_i s_i$ are independent, then random variables $\{s_i\}$ for which $\alpha_i \beta_i \neq 0$ are complex Gaussian.

A. Superadditive Functional and Contrast

Definition 4: A real functional of the distribution of a circular complex random variable z , denoted as $Q(z)$, is said to be of class II (of a distribution) if it is

- (i) *translation invariant*, in the sense that $Q(z + \alpha) = Q(z)$ for any complex number α ;
- (ii) *scale equivariant*, in the sense that $Q(\alpha z) = |\alpha| Q(z)$ for any nonzero complex number α ; and
- (iii) *rotation invariant*, in the sense that $Q(z) = Q(ze^{j\alpha})$ for any nonzero real number α .

Definition 5: The functional Q is *superadditive* if

$$Q^2(z_1 + z_2) \geq Q^2(z_1) + Q^2(z_2) \quad (16)$$

for any two mutually independent circular complex random variables z_1 and z_2 ; in contrast, Q is said to be *k-subadditive* ($0 < k \in \mathbb{R}$) if $Q^k(z_1 + z_2) \leq Q^k(z_1) + Q^k(z_2)$.

Theorem 2: If Q is a superadditive functional of class II, then a contrast function for separating instantaneous linear mixtures of circular complex sources has the following form:

$$J(\mathbf{W}) = \sum_{i=1}^n \log Q(y_i) - \log |\det(\mathbf{W})|, \quad (17)$$

which is discriminating if $Q(s_j) > 0$ for every independent source s_j and if the inequality is strict for any pair of nonzero multiples of two distinct complex sources.

Note that when \mathbf{W} is a unitary matrix (i.e., $\mathbf{W}\mathbf{W}^H = \mathbf{I}$), then (17) is simplified to $J(\mathbf{W}) = \sum_{i=1}^n \log Q(y_i)$.

Proof: The proof follows almost the same as the real-valued case [15] and is repeated here for completeness. First, let $\mathbf{y} = \mathbf{W}(\mathbf{A}\mathbf{s}) = \mathbf{R}\mathbf{s}$, where $\mathbf{R} = [r_{ij}]_{n \times n}$ is an $n \times n$ complex-valued matrix; in the scalar form, we have $y_i = \sum_{j=1}^n r_{ij} s_j$. In light of the scale equivariance and superadditivity of Q , we have

$$Q^2(y_i) \geq \sum_{j=1}^n |r_{ij}|^2 Q^2(s_j) \quad (i = 1, \dots, n). \quad (18)$$

Let $\mathbf{\Lambda} = \mathbf{R} \text{diag}\{Q^2(s_1), \dots, Q^2(s_n)\} \mathbf{R}^H$, then $Q^2(y_i)$ is the i th diagonal entry of the *Hermitian* matrix $\mathbf{\Lambda}$. If $Q(s_j) > 0$, then $\mathbf{\Lambda}$ is a positive definite matrix. In light of the *Hadamard inequality*, the product of the diagonal elements of a positive definite matrix is lower bounded by its determinant (with equality of bound if and only if the matrix is diagonal), then equation (17) is bounded below by $\sum_{j=1}^n \log Q(s_j) + \log |\det(\mathbf{R})| - \log |\det(\mathbf{W})|$; if $\mathbf{\Lambda}$ is singular, then the lower bound approaches $-\infty$; Since $\det(\mathbf{R})/\det(\mathbf{W}) = \det(\mathbf{A})$, the bound is attained when \mathbf{R} equals to the product of a permutation matrix \mathbf{P} and a complex diagonal matrix \mathbf{D} , namely, $\mathbf{R} = \mathbf{P}\mathbf{D}$. Second, to show the contrast function is discriminating, we need to prove the lower bound is obtained when $\mathbf{R} = \mathbf{P}\mathbf{D}$. This can be seen that (17) attains its minimum only if (18) achieves equality. For the i th equality, we would have

$$Q^2(r_{ij} s_j + r_{ik} s_k) = Q^2(r_{ij} s_j) + Q^2(r_{ik} s_k) \quad (j \neq k),$$

which nevertheless contradicts the assumption unless r_{ij} or r_{ik} is zero. Therefore, for each index i , there is at most one index j for which $r_{ij} \neq 0$. Given the assumptions that $Q(y_i) > 0$ and matrix \mathbf{R} is nonsingular, each row and column of the matrix \mathbf{R} must merely contain one nonzero complex numbers; thus \mathbf{R} can be written as a product of permutation matrix \mathbf{P} and a complex diagonal matrix \mathbf{D} . \square

Corollary 1: If Q is a functional of class II and if

$$Q(z_1 + z_2) = Q(z_1) + Q(z_2) \quad (19)$$

holds for two mutually independent circular complex variables z_1 and z_2 ; then (17) is an effective contrast function for separating instantaneous linear mixtures of independent circular complex sources; it is discriminating if $Q(s_j) > 0$ for every independent source s_j .

Proof: Notably the property (19) is much stronger than the superadditivity. If Q is positive, from (19) it follows that

$$\begin{aligned} Q^2(z_1 + z_2) &= Q^2(z_1) + Q^2(z_2) + 2Q(z_1)Q(z_2) \\ &\geq Q^2(z_1) + Q^2(z_2), \end{aligned}$$

and the equality holds if and only if $Q(z_1) = Q(z_2) = 0$. The rest of the proof is similar to that of Theorem 2. \square

Theorem 2 and Corollary 1 essentially provide *sufficient* criteria to design contrast functions, therefore finding superadditive functional of class II is the key. However, unlike real-valued case, most contrast functions (e.g., [4], [15]) for real ICA are neither superadditive nor subadditive in the general complex ICA setup. For instance, the k -order cumulant statistic (which is subadditive in real domain for $k > 2$) would *not* satisfy $\text{cum}_k(z_1 + z_2) = \text{cum}_k(z_1) + \text{cum}_k(z_2)$ for two independent complex random variables z_1 and z_2 . This is due to the phase dependence of sum of complex numbers. To illustrate this, let $z_1 = |z_1|e^{j\theta_1}$, $z_2 = |z_2|e^{j\theta_2}$ ($\theta_1, \theta_2 \in \mathbb{R}$) be two complex random variables, then their sum is represented as

$$\begin{aligned} z_1 + z_2 &= |z_1|e^{j\theta_1} + |z_2|e^{j\theta_2} \\ &= \sqrt{|z_1|^2 + |z_2|^2 + 2|z_1||z_2|\cos(\theta_1 - \theta_2)}e^{j\theta}, \end{aligned}$$

where $\theta = \arctan \frac{z_1 \sin \theta_1 + z_2 \sin \theta_2}{z_1 \cos \theta_1 + z_2 \cos \theta_2}$. Obviously, the modulus of the sum depends on $\theta_1 - \theta_2$, and we have $||z_1| - |z_2|| \leq |z_1 + z_2| \leq |z_1| + |z_2|$.

Theorem 3: If Q is a superadditive (or 2-subadditive) functional of class II in real domain; let $z = |z|e^{j\theta}$ ($\theta \in \mathbb{R}$), then $\overline{Q}(z) = Q(|z|) = \sqrt{Q^2(|z|\cos\theta) + Q^2(|z|\sin\theta)}$ is also a superadditive (or 2-subadditive) functional of class II in complex domain.

Proof: Consider superadditivity first. Let $z_1 = |z_1|e^{j\theta_1}$ and $z_2 = |z_2|e^{j\theta_2}$ be two independent circular complex random variables; let $z = |z|e^{j\theta} = z_1 + z_2$, then $z_{\text{Re}} = z_1^{\text{Re}} + z_2^{\text{Re}}$ and $z_{\text{Im}} = z_1^{\text{Im}} + z_2^{\text{Im}}$; because of the superadditivity of Q in real domain, we have

$$\begin{aligned} Q^2(|z|\cos\theta) &\geq Q^2(|z_1|\cos\theta_1) + Q^2(|z_2|\cos\theta_2) \\ Q^2(|z|\sin\theta) &\geq Q^2(|z_1|\sin\theta_1) + Q^2(|z_2|\sin\theta_2). \end{aligned}$$

Summing up the above two inequalities and because of scale equivariance, we obtain

$$\begin{aligned} \overline{Q}^2(z) &= Q^2(|z|) = Q^2(|z|\cos\theta) + Q^2(|z|\sin\theta) \\ &\geq Q^2(|z_1|) + Q^2(|z_2|) = \overline{Q}^2(z_1) + \overline{Q}^2(z_2), \end{aligned} \quad (20)$$

which satisfies the Definition 5 and completes the proof. Similarly, we can also prove 2-subadditivity. \square

Hence, for circular complex variables, if we can find a functional of class II, while its real counterpart is superadditive, with Theorem 3 we may always find a superadditive functional of class II in complex domain.

Lemma 1: If $Q(z)$ ($z \in \mathbb{C}$) is a real-valued function of class II in complex domain; let $\tilde{z} = [z_{\text{Re}}, z_{\text{Im}}]$, then Q is also a function of class II in real domain, with argument of $\tilde{z} \in \mathbb{R}^2$; specifically, it satisfies (i) translation invariance: $Q(\alpha + \tilde{z}) = Q(\tilde{z})$ ($\forall \alpha \in \mathbb{R}^2$); and (ii) scale equivariance: $Q(\alpha\tilde{z}) = |\alpha|Q(\tilde{z})$ ($\forall \alpha \in \mathbb{R}$).

Proof: The proof is straightforward. The first property is obvious. The second property can be verified by letting $\alpha = |\alpha|e^{j0}$ (i.e., zero phase). Notably the converse is not true. \square

Similar to the real-valued ICA [15], we can also construct contrast functions based on subadditive functional in the complex ICA setting; specifically, a theorem follows in the below.

Theorem 4: Given a unitary mixing matrix, if Q is a k -subadditive functional of class II for some $k \geq 2$, then $-\sum_{i=1}^n Q^k(y_i)$ and $-\sum_{i=1}^n Q^{2k}(y_i)$ are contrast functions for separating an instantaneous mixture of independent circular complex sources when the demixing matrix is also unitary; the contrasts are discriminating if $k > 2$ and for sources $\{s_j\}$, there is at most one index j for which $Q(s_j) = 0$.

Proof: The proof is close to that of [15] and only sketched here. When $k = 2$ and $\mathbf{R} = \mathbf{W}\mathbf{A}$ is unitary, by subadditivity of Q , we obtain $Q^2(y_i) \leq \sum_{j=1}^n |r_{ij}|^2 Q^2(s_j) \leq \sum_{j=1}^n Q^2(s_j)$ because of $|r_{ij}| \leq \sum_{k=1}^n |r_{ik}|^2 = 1$; when $k > 2$, we have $Q^k(y_i) \leq \sum_{j=1}^n |r_{ij}|^k Q^k(s_j) \leq \sum_{j=1}^n Q^k(s_j)$. It then follows that $-\sum_{i=1}^n Q^k(y_i) \geq -\sum_{j=1}^n Q^k(s_j)$, hence $-\sum_{i=1}^n Q^k(y_i)$ is a contrast function. The rest of the proof is omitted due to lack of space. \square

Remark: The constraint of unitary mixing matrix can be satisfied by a whitening procedure with EVD: $\mathbb{E}[\mathbf{x}\mathbf{x}^H] = \mathbf{A}\mathbb{E}[\mathbf{s}\mathbf{s}^H]\mathbf{A}^H = \mathbf{U}\Sigma\mathbf{U}^H$, where Σ is a diagonal matrix with nonnegative real-valued entries, and \mathbf{U} is a unitary matrix. Define a linear transformation $\mathbf{z} = \mathbf{D}^{-1/2}\mathbf{U}^H\mathbf{x} = \mathbf{D}^{-1/2}\mathbf{U}^H\mathbf{A}\mathbf{s} \equiv \tilde{\mathbf{A}}\mathbf{s}$, where $\tilde{\mathbf{A}} = \mathbf{D}^{-1/2}\mathbf{U}^H\mathbf{A}$ denotes a new mixing matrix such that $\mathbb{E}[\mathbf{z}\mathbf{z}^H] = \tilde{\mathbf{A}}\mathbb{E}[\mathbf{s}\mathbf{s}^H]\tilde{\mathbf{A}}^H = \mathbf{I}$; the new mixing matrix $\tilde{\mathbf{A}}$ becomes unitary if $\mathbb{E}[\mathbf{s}\mathbf{s}^H] = \mathbf{I}$.

B. Examples of Contrast Functions

1) *Range Function:* In [15] and [21], range function was used as a contrast function for real-valued ICA, we may also generalize its use to complex domain. Assume all complex variables are distributed within bounded support circles, i.e., the probability that the complex variables is inside the bounded or finite support circle is nonnegative, and zero elsewhere. Correspondingly, the range of a complex

number z , denoted by $R(z)$, is defined by the radius of its support circle in the complex plane:

$$R(z) = d, \quad \text{where} \quad \{p(z) \geq 0 \mid |z - z_0| \leq d\} \quad (21)$$

where z_0 denotes the center of the support circle, and the positive scalar $d \in \mathbb{R}^+$ denotes the radius.

Theorem 5: If two circular complex bounded random variables z_1 and z_2 are independent, then their range satisfies the linear superposition principle

$$R(z_1 + z_2) = R(z_1) + R(z_2). \quad (22)$$

Proof: The proof follows the simple fact that $|z_1 + z_2| \leq |z_1| + |z_2|$ and range is a maximum operation. \square

Corollary 2: If $z = \alpha z_1 + \beta z_2$, where z_1 and z_2 are two independent circular complex bounded random variables, and $\alpha, \beta \in \mathbb{C}$, then

$$R(z) = |\alpha| \cdot R(z_1) + |\beta| \cdot R(z_2). \quad (23)$$

The proof follows directly from Theorem 5 and the fact that the radius of the support circle of αz_1 (or βz_2) is just $|\alpha|$ (or $|\beta|$) multiple of the radius of z_1 (or z_2), which is independent on the phase of α (or β).

Let $Q(y_i) = R(y_i)$, then (17) may be rewritten as

$$J(\mathbf{W}) = \sum_{i=1}^n \log \left(\sum_{j=1}^n |r_{ij}| R(s_j) \right) - \log |\det(\mathbf{W})|. \quad (24)$$

With little modification, the contrast function may be also used for finite circular complex sources.

Definition 6: Let z be a circular complex random variable with unbounded support, $\{z : |z| \leq +\infty\}$; z is said to be finite if $\int_0^d p_z(|z|) d|z| \approx 1$ ($0 < d < +\infty$). Correspondingly, we can view d as the radius parameter of the finite variable z .

2) **Shannon Entropy Function:** A well-known information-theoretic contrast function is the marginal Shannon entropy, $H(y_i)$. Let

$$\overline{Q}(y_i) = Q(|y_i|) = \exp(H(|y_i|)), \quad (25)$$

then from *entropy power inequality*, it is known that (25) is a superadditive functional of class II in real domain. From Theorem 3 we can also construct a superadditive functional of class II for circular complex variables. It is observed that plugging (25) into (17) yields (6) except for a constant difference. Given finite samples of a complex random variable $\{y_i\}$, one can obtain the modulus $\{|y_i|\}$ and estimate their pdfs/entropies using the conventional real-valued pdf/entropy estimators, such as the kernel estimator or spacing estimator [16],[17].

3) **Rényi Entropy Function:** In addition to Shannon entropy, we can also define the generalized k -order Rényi entropy ($0 < k \in \mathbb{R}$):

$$H_k(y_i) = \frac{1}{1-k} \log \left(\int p(y_i)^k dy_i \right). \quad (26)$$

When the limit $k \rightarrow 1$ is taken, Rényi entropy reduces to the standard Shannon entropy. When $k = 2$, Rényi entropy

of order 2 is often called *extension entropy*

$$H_2(y_i) = -\log \left(\int p(y_i)^2 dy_i \right).$$

By virtue of the *Jensen inequality*, we have $H_2(y_i) \leq H(y_i)$. In general, Rényi entropy is a non-increasing function in the sense that $H_k(y_i) \geq H_r(y_i)$ for any $r > k$.

Lemma 2: For circular complex variable $\{z : p(z) = g(|z|)\}$, the exponent of quadratic Rényi entropy, $\exp(H_2(z))$, is a functional of class II.

Proof: First, it is easy to see $H_2(z) = H_2(z + \alpha)$ for any complex number α . Next, we have

$$\begin{aligned} H_2(\alpha z) &= -\log \left(\int g(|\alpha z|)^2 d|\alpha z| \right) \\ &= -\log \left(\int g(|z|)^2 d|z| \right) + \log |\alpha|. \end{aligned}$$

and then $\exp(H_2(\alpha z)) = |\alpha| \exp(H_2(z))$. Similarly, we can prove $\exp(H_2(z))$ is rotation invariant. \square

Specifically, by letting $\overline{Q}(y_i) = Q(|y_i|) = \exp(H_2(|y_i|)) = \frac{1}{\int p(|y_i|)^2 dy_i}$, we obtain the generalized 2-Rényi entropy power. By introducing generalized Gaussian densities, generalized power inequalities can be similarly derived [26,27]; then \overline{Q} is a superadditive functional of class II. Given a finite number of samples of $\{|y_i|\}$, an efficient quadratic Rényi entropy kernel estimator is available [18].

4) **Fisher Information Function:** Let us further consider the functional of Fisher information

$$\mathbf{G} = \mathbb{E}[\psi(z)^2] = \mathbb{E} \left[\left(\frac{d \log p(z)}{dz} \right)^2 \right], \quad (27)$$

where $\psi(z) = \frac{d \log p(z)}{dz} = \frac{p'(z)}{p(z)}$ is called the score function. If we let $Q(z) = \mathbf{G}^{-1/2}$, it can be proved that Q is a functional of class II that is superadditive in real domain [15]. For circular complex variables, we obtain

$$\overline{Q}(z) = \mathbb{E} \left[\left(\frac{p'(|z|)}{p(|z|)} \right)^2 \right]^{-1/2}.$$

In parallel with the k -order Rényi entropy, the k -score function can be defined as $\psi_k(z) = \frac{p'(z)}{p(z)^{2-k}}$, and the k -Fisher information matrix can also be defined as [26]:

$$\mathbf{G}_k(z) = \frac{\int p(z) (\psi_k(z))^2 dz}{\int p(z)^k dz}. \quad (28)$$

5) **Determinant Function:** Let $\omega \in \mathbb{C}$ (which is deterministic) and $z \in \mathbb{C}$ (which is random) denote two complex variables, and denote

$$\Omega = \begin{bmatrix} \omega_{\text{Re}} & -\omega_{\text{Im}} \\ \omega_{\text{Im}} & \omega_{\text{Re}} \end{bmatrix}, \quad \tilde{\mathbf{z}} = \begin{bmatrix} z_{\text{Re}} \\ z_{\text{Im}} \end{bmatrix},$$

and $\Sigma_z = \text{cov}[\tilde{\mathbf{z}}]$. Let $z' = \omega z$, then

$$\tilde{\mathbf{z}}' = \begin{bmatrix} z'_{\text{Re}} \\ z'_{\text{Im}} \end{bmatrix} = \Omega \tilde{\mathbf{z}} = \begin{bmatrix} \omega_{\text{Re}} z_{\text{Re}} - \omega_{\text{Im}} z_{\text{Im}} \\ \omega_{\text{Im}} z_{\text{Re}} + \omega_{\text{Re}} z_{\text{Im}} \end{bmatrix}.$$

If we let $Q(z) = \det(\Sigma_z)^{1/2} \equiv \det(\text{cov}[\tilde{\mathbf{z}}])^{1/2}$, then

$$\begin{aligned} Q(\omega z) &= \det(\text{cov}[\tilde{\mathbf{z}}'])^{1/2} = \det(\text{cov}[\Omega\tilde{\mathbf{z}}])^{1/2} \\ &= \det(\Omega\text{cov}[\tilde{\mathbf{z}}])^{1/2} = \det(\Omega)^{1/2}\det(\text{cov}[\tilde{\mathbf{z}}])^{1/2} \\ &= (\omega_{\text{Re}}^2 + \omega_{\text{Im}}^2)^{1/2}\det(\Sigma_z)^{1/2} = |\omega|\det(\Sigma_z)^{1/2}. \end{aligned}$$

Therefore $Q(z)$ is scale equivariant. It is also easy to prove $Q(z)$ is translation and rotation invariant; thus $Q(z)$ is a functional of class II. If z is circular, then it is also straightforward to verify $Q(z)$ is superadditive.

Correspondingly, let $Q(y_i) = \det(\text{cov}[y_i^{\text{Re}}, y_i^{\text{Im}}])^{1/2}$, we can design the following contrast function

$$J(\mathbf{W}) = \sum_{i=1}^n \log \left(\sum_{j=1}^n |r_{ij}| \det(\text{cov}[s_j^{\text{Re}}, s_j^{\text{Im}}])^{1/2} \right), \quad (29)$$

where $\log |\det(\mathbf{W})| = 0$ holds for a unitary matrix \mathbf{W} . Note that if s_j^{Re} is uncorrelated with s_j^{Im} , then their covariance matrix is diagonal, and $\det(\text{cov}[s_j^{\text{Re}}, s_j^{\text{Im}}]) = \text{var}[s_j^{\text{Re}}]\text{var}[s_j^{\text{Im}}]$.

6) *Modified Kurtosis Function*: Let us further consider a modified kurtosis function as follows

$$Q(y_i) = \left(\mathbb{E} \left[|y_i - \mathbb{E}[y_i]|^4 \right] \right)^{1/4}. \quad (30)$$

It is easy to prove that (30) is a functional of class II. In the meantime, Q is superadditive in real domain if the independent real variables (e.g., $x, y \in \mathbb{R}$) are sub-Gaussian [15], such that $Q^4(x+y) \geq Q^4(x) + Q^4(y)$.

Assuming $\mathbb{E}[\mathbf{s}] = \mathbf{0}$ and $\mathbb{E}[y_i] = 0$, $|\det(\mathbf{W})| = 1$, and denoting $\gamma_j = \mathbb{E}[|s_j|^4]$, then the contrast function can be rewritten as

$$J(\mathbf{R}) = \sum_{i=1}^n \frac{1}{4} \log \left(\sum_{j=1}^n |r_{ij}|^4 \gamma_j \right). \quad (31)$$

V. CASE STUDY: TWO-BY-TWO MIXTURE

In this section, we use a simple (2-by-2 mixing) example to illustrate the contrast function analysis in the context of local/global minima or maxima. Specifically, we consider the range contrast function defined in (24). For discussion simplicity, we assume the sources, mixtures, and outputs are all whitened such that $\mathbb{E}[\mathbf{s}\mathbf{s}^H] = \mathbb{E}[\mathbf{x}\mathbf{x}^H] = \mathbb{E}[\mathbf{y}\mathbf{y}^H] = \mathbf{I}$, which implies that \mathbf{W} and \mathbf{R} are both unitary matrices. In general, the unitary matrix \mathbf{R} can be written with the following general form:

$$\mathbf{R} = \begin{bmatrix} e^{j\beta_1} \cos \theta & e^{j\beta_2} \sin \theta \\ -e^{-j\beta_2} \sin \theta & e^{-j\beta_1} \cos \theta \end{bmatrix} \triangleq \begin{bmatrix} c & s \\ -s^* & c^* \end{bmatrix}, \quad (32)$$

with $\beta_1, \beta_2 \in \mathbb{R}$ and $\theta \in [0, 2\pi)$; $\det(\mathbf{R}) = |c|^2 + |s|^2 = 1$. Because $\log |\det(\mathbf{W})| = 0$, equation (24) may be written in the form of $J(\mathbf{R})$:

$$J(\mathbf{R}) = \sum_{i=1}^2 \log \left(\sum_{j=1}^2 |r_{ij}| R(s_j) \right). \quad (33)$$

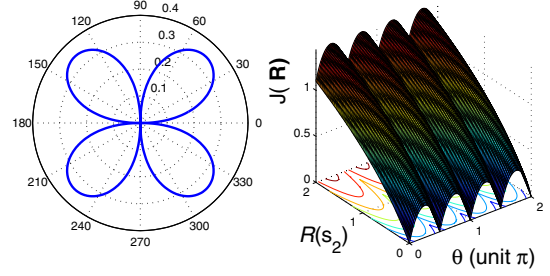


Fig. 2. Figural illustration of local maxima and global minima of $J(\mathbf{R})$ w.r.t. $\theta \in [0, 2\pi)$, where we assume the source range $R(s_1) = R(s_2) = 1$ in the left plot and $R(s_1) = 1$ in the right plot.

Plugging (32) into (33), and differentiating $J(\mathbf{R})$ w.r.t. θ and letting $\frac{dJ(\mathbf{R})}{d\theta} = 0$, we obtain

$$\left(R^2(s_1) + R^2(s_2) \right) \left(\text{sign}(\sin 2\theta) (\cos^2 \theta - \sin^2 \theta) \right) = 0.$$

Since $R(s_1)$ and $R(s_2)$ are positive, therefore the stationary point is the solution of $\text{sign}(\sin 2\theta) (\cos^2 \theta - \sin^2 \theta) = 0$, then the solutions are either $\theta = \frac{k\pi}{2}$ ($k = 0, 1, 2, 3$), or $\theta = \frac{\pi}{4} + \frac{k\pi}{2}$ ($k = 0, 1, 2, 3$). However, it is noted that the function $J(\mathbf{R})$ at stationary points $\frac{k\pi}{2}$ is not continuous, and therefore is non-differentiable.

Calculating the Hessian (for $\theta \neq k\pi/2$) would further yield

$$\begin{aligned} \frac{d^2 J(\mathbf{R})}{d\theta^2} &= \frac{4\ell_1 \left(2\delta(\sin 2\theta) \cos^2 2\theta - \text{sign}(\sin 2\theta) \sin 2\theta \right)}{\left(|\sin 2\theta| \ell_1 + (\ell_2 - \ell_1) \right)} \\ &\quad - \frac{4\ell_1^2 \cos^2 2\theta}{\left(|\sin 2\theta| \ell_1 + (\ell_2 - \ell_1) \right)^2}, \end{aligned}$$

where we denote $\ell_1 = R^2(s_1) + R^2(s_2)$, $\ell_2 = (R(s_1) + R(s_2))^2$, and $R(s_1)R(s_2) = \frac{\ell_2 - \ell_1}{2}$. It then concludes

- When $\theta = \frac{\pi}{4} + \frac{k\pi}{2}$, $\sin 2\theta = \pm 1$ and $\cos 2\theta = 0$, and the Hessian value at the stationary points is $\frac{-4\ell_1}{\ell_2} < 0$, thus the stationary points are only local maxima.
- When $\theta = \frac{k\pi}{2}$, because there are no other stationary points for $J(\mathbf{R})$ to be minimized, it can be inferred that they all correspond to the global minima.

Figure 2 shows a graphical illustration of all local maxima and global minima of (33) within the region $[0, 2\pi)$. The above analysis procedure is quite general, similar analysis can also be conducted for other contrast functions, such as equations (29) and (31).

VI. DISCUSSION

In this paper, we have focused on the design criteria of contrast functions for complex ICA, especially for circular

complex sources. Obviously, there are still many important theoretical issues that require further investigations. A number of them are listed here.

- As reported in the real ICA, spurious local minima exist for information-theoretic criteria (such as mutual information or negentropy) as well as cumulant-based contrast (such as kurtosis) [22]-[24], it is interesting to examine such phenomena in complex ICA.
- Typical optimization algorithms for ICA use local gradient search, with or without unitary constraint [25] (depending if \mathbf{W} is a unitary matrix). The $n \times n$ unitary matrices form a unitary group $\mathcal{U}(n)$, which is a Lie group of dimension n^2 . Such a special structure may require special design of the optimization algorithm.
- Although we only discuss simultaneous separation approach of complex ICA in this paper, it would be also interesting to investigate the contrast function for the sequential extraction (i.e., deflation) approach, which can be viewed as a degenerate case of the former approach. In both cases, stability analysis of the algorithm is important.

VII. CONCLUDING REMARKS

Complex-valued ICA problems occur in many practical applications, such as separation of fMRI images or EEG [7, 13], speech processing in frequency domain [14], communications and array signal processing [5, 12]. Recent years have witnessed many developments of complex ICA algorithms [6]-[14]. However, the theory of complex ICA is still not as well understood as its real counterpart, especially for simultaneous separation. Essentially, designing effective contrast functions is the key to the solution. In this paper, a number of contrast functions have been studied for the circular complex sources, and we overview and extend some established work and derive several new results. The current paper is purely theoretically oriented; practical implementations and design of efficient algorithms are the topics of future investigation.

APPENDIX

For a real-valued random variable u (which can be either the real/imaginary component, or its modulus, of a complex number), we may obtain its up-to-fourth-order Gram-Charlier expansion $p(u) \approx \mathcal{N}(u) \left(1 + \frac{\kappa_3}{6} \mathcal{H}_3(u) + \frac{\kappa_4}{24} \mathcal{H}_4(u) \right)$, where $\mathcal{H}_k(u)$ denotes the k -order Chebyshev-Hermite polynomial, κ_k denotes the k -order cumulant of u , and $\mathcal{N}(u)$ denotes the Gaussian probability density as $\mathcal{N}(u) = \frac{1}{\sqrt{2\pi}} \exp(-u^2/2)$. Upon some approximation [2], we may derive

$$H(u) \approx \frac{1}{2} \log(2\pi e) - \frac{\kappa_3^2}{12} - \frac{\kappa_4^2}{48} + \frac{3}{8} \kappa_3^2 \kappa_4 + \frac{\kappa_4^3}{16},$$

and the relevant partial derivatives of (12): $\frac{\partial \log |\det(\mathbf{W})|}{\partial w_{ij}^*} = (\mathbf{W}^{-H})_{ij}$, $\frac{\partial H(y_i^{\text{Im}})}{\partial w_{ij}^*} \approx \left(-\frac{3}{2} (y_i^{\text{Im}})^5 - \frac{239}{12} (y_i^{\text{Im}})^7 - \frac{15}{4} (y_i^{\text{Im}})^9 + \frac{3}{2} (y_i^{\text{Im}})^{11} \right) x_j$, and similarly for $\frac{\partial H(e_i)}{\partial w_{ij}^*}$ (omitted here due to its lengthy expression).

REFERENCES

- [1] A. Hyvärinen, J. Karhunen and E. Oja, *Independent Component Analysis*. New York: Wiley, 2001.
- [2] A. Cichocki and S. Amari, *Adaptive Blind Signal and Image Processing*. New York: Wiley, 2002.
- [3] P. Comon, "Independent component analysis, a new concept?" *Signal Processing*, vol. 36, no. 3, pp. 287–314, 1994.
- [4] J.-F. Cardoso, "Blind signal separation: Statistical principles," *Proc. IEEE*, vol. 90, pp. 2009–2026, 1998.
- [5] J.-F. Cardoso, "An efficient technique for the blind separation of complex sources," *Proc. Higher-Order Statistics (HOS'93)*, pp. 275–279, South Lake Tahoe, CA, 1993.
- [6] E. Bingham and A. Hyvärinen, "A fast fixed-point algorithm for independent component analysis of complex valued signals," *International Journal of Neural Systems*, vol. 10, no. 1, pp. 1–8, 2000.
- [7] V. Carhoun and T. Adali, "Complex Infomax: convergence and approximation of Infomax with complex nonlinearities," *Proc. IEEE Workshop on Neural Networks for Signal Processing*, Switzerland, 2002.
- [8] T. Adali, T. Kim and V. Calhoun, "Independent component analysis by complex nonlinearities," *Proc. ICASSP2005*, V-525-528, Philadelphia, 2005.
- [9] J.-F. Cardoso and T. Adali, "The maximum likelihood approach to complex ICA," *Proc. ICASSP2006*, to appear.
- [10] J. Eriksson and V. Koivunen, "Complex random vectors and ICA models: Identifiability, uniqueness and separability," *IEEE Trans. Info. Theory*, vol. 52, no. 3, 2006, pp. 1017–1029.
- [11] J. Eriksson, A.-M. Seppola and V. Koivunen, "Complex ICA for circular and non-circular sources," *Proc. the 13th European Signal Processing Conference (EUSIPCO'2005)*, Turkey, Sept., 2005.
- [12] S. Fiori, "Extended Hebbian learning for blind separation of complex-valued sources," *IEEE Trans. Circuits and Systems II*, vol. 50, no. 4, pp. 195–202, 2003.
- [13] J. Anemuller, T. J. Sejnowski, and S. Makeig, "Complex independent component analysis of frequency-domain electroencephalographic data," *Neural Networks*, vol. 16, no. 1311–1323, 2003.
- [14] P. Smaragdis, "Blind separation of convolved mixtures in the frequency domain," *Neurocomputing*, vol. 22, pp. 21–34, 1998.
- [15] D. T. Pham, "Contrast functions for ICA and sources separation," Tech. Rep., May, 2001. Available online at <http://www-lmc.imag.fr/lmc-sms/Dinh-Tuan.Pham/publics-tries.html>
- [16] D. T. Pham, "Fast algorithm for estimating mutual information, entropies and score functions," *Proc. ICA2003*, Nara, Japan, 2003.
- [17] E. G. Miller and J. W. Fisher II, "ICA using spacings estimates of entropy," *J. Machine Learning Res.*, vol. 4, pp. 1271–1295, 2003.
- [18] K. Hild, D. Erdogmus, and J. Principe, "Blind source separation using Rényi's mutual information," *IEEE Signal Processing Letters*, vol. 8, no. 6, pp. 174–176, 2001.
- [19] P. J. Huber, "Projection pursuit," *Ann. Statist.*, vol. 13, no. 2, pp. 435–475, 1985.
- [20] S. A. Cruces-Alvarez, A. Cichocki, and S. Amari, "From blind signal extraction to blind instantaneous signal separation: Criteria, algorithms, and stability," *IEEE Trans. Neural Networks*, vol. 15, no. 4, pp. 859–872, 2005.
- [21] F. Vrins, M. Verleysen, and C. Jutten, "SWM: A class of convex contrasts for source separation," *Proc. ICASSP2005*, V-161-164, Philadelphia, 2005.
- [22] J. Ma, Z. Chen and S.-I. Amari, "Analysis of feasible solutions of the ICA problem under the one-bit-matching condition," *Proc. ICA2006*, Lecture Notes in Computer Science, vol. 3889, pp. 838–845, 2006.
- [23] F. Vrins and M. Verleysen, "Information theoretic vs cumulant-based contrast for multimodal source separation," *IEEE Signal Processing Letters*, vol. 12, no. 3, pp. 190–193, 2005.
- [24] D. T. Pham and F. Vrins, "Local minima of information-theoretic criteria in blind source separation," *IEEE Signal Processing Letters*, vol. 12, no. 11, pp. 788–791, 2005.
- [25] Y. Nishimori and S. Akaho, "Learning algorithms utilizing quasi-geodesic flows on the Stiefel manifold," *Neurocomputing*, vol. 67, pp. 106–135, 2005.
- [26] O. Johnson and C. Vignat, "Some results concerning maximum Rényi entropy distributions," paper preprint, math.PR/0507400, July, 2005
- [27] E. Lutwak, D. Yang, and G. Zhang, "Cramér-Rao and moment-entropy inequalities for Rényi entropy and generalized Fisher information," *IEEE Trans. Information Theory*, vol. 51, No. 2, pp. 473–478, 2005.