A PRACTICAL FORMULATION FOR COMPUTATION OF COMPLEX GRADIENTS AND ITS APPLICATION TO MAXIMUM LIKELIHOOD ICA

Tülay Adalı and Hualiang Li

University of Maryland Baltimore County
Baltimore, MD 21250, USA

ABSTRACT

We introduce a framework for complex-valued signal processing such that all computations can be directly carried out in the complex domain. The framework, based on an elegant result due to Brandwood, allows for easy derivation of many complex-valued algorithms and their efficient analyses. We demonstrate its application to derivation of relative gradient updates for independent component analysis using maximum likelihood and discuss the selection of score functions within this framework.

***Index Terms*—** Maximum likelihood estimation, optimization methods, signal analysis.

1. INTRODUCTION

Complex-valued signals arise frequently in applications as diverse as communications, radar, and biomedicine, as most practical modulation formats are of complex type and applications such as radar and magnetic resonance imaging lead to data that are inherently complex valued. When the processing has to be done in a transform domain such as Fourier or complex wavelet, again the data are complex valued. In order to perform independent component analysis (ICA) of complex-valued data there are a number of options. Algorithms such as joint approximate diagonallization of eigenmatrices (JADE) [11] or those using second order statistics [13] achieve ICA without the need to use nonlinear functions in the algorithm. The second-order complex ICA algorithm, strongly uncorrelating transform (SUT), though is very efficient, requires the signals to be noncircular and a second algorithm should be utilized after its application as a preprocessing step when the sources happen to be circular [14], and JADE's performance–as well as that of any algorithm that uses joint diagonalizations—suffers when the number of sources increase (see *e.g.* [18]). On the other hand, nonlinear ICA approaches such as maximum likelihood (ML) [19], information-maximization (Infomax) [8], nonlinear decorrelations [12], and maximization of nongaussianity [15], which are all intimately related to each other, generate higher-order statistics implicitly using nonlinear functions, and thus present an attractive alternative for performing ICA. A number of comparison studies have demonstrated their desirable performance over other ICA algorithms such as JADE and second-order algorithms.

In this paper, we use an elegant result due to Brandwood [6] to introduce a framework for complex-valued signal processing such that all computations can be carried out in the complex domain. Thus, the need for simplifying assumptions such as circularity of sources are largely eliminated both in the derivation and the analysis of the algorithms. We show how this framework can be utilized to derive the update rules for ML ICA in a very straightforward manner and allows for a new view of the selection of score functions enabling working entirely in the complex domain.

This work was supported by the NSF grants NSF-CCF 0635129 and NSF-IIS 0612076.

2. BRANDWOOD'S ANALYTICITY CONDITION AND COMPLEX GRADIENTS

To introduce Brandwood's result [6], which plays a key role in our development, we define g: $\mathbb{C} \times \mathbb{C} \to \mathbb{C}$ as a function of a complex variable z and its conjugate z^*. If treating z (resp. z^*) as a constant, g is analytic on z^* (resp. z), then we say that g satisfies *Brandwood's analyticity condition* (BAC). This concept can be similarly extended to vector and matrix quantities. Because our main interest is in functions g that are *cost* functions, we consider the more special case of g: $\mathbb{C} \times \mathbb{C} \to \mathbb{R}$ and state the main result of [6] for these class of functions as:

Theorem: Let f: $\mathbb{R} \times \mathbb{R} \to \mathbb{R}$ be a function of real variables x and y such that $g(z, z^*) = f(x, y)$, where $z = x + jy$ and that g satisfies the BAC. Then, the partial derivative $\partial g/\partial z$ (treating z^* as a constant in g) gives the same result as $(\partial f/\partial x - j\partial f/\partial y)/2$ on substituting for z. Similarly, $\partial g/\partial z^* = (\partial f/\partial x + j\partial f/\partial y)/2$.

Thus, when evaluating the gradient of functions, we can directly compute the derivatives with respect to the complex argument, rather than calculating individual real-valued gradients as typically performed in literature (see *e.g.*, [14], [16], [20]). It is also stated that a necessary and sufficient condition for f to have a stationary point is that $\partial g/\partial z = 0$ or $\partial g/\partial z^* = 0$.

The main result stated above can be easily extended to vector and matrix gradients by defining the scalar inner product between two matrices $\mathbf{W}$ and $\mathbf{V}$ as $\langle \mathbf{W}, \mathbf{V}\rangle = \mathrm{Trace}(\mathbf{W}^H\mathbf{V})$ so that $\langle \mathbf{W}, \mathbf{W}\rangle = \|\mathbf{W}\|^2_{\mathrm{Fro}}$, where the subscript denotes the Frobenius norm. For vectors, the definition simplifies to $\langle \mathbf{w}, \mathbf{v}\rangle = \mathbf{w}^H\mathbf{v}$. We can define the gradient vector $\nabla_{\mathbf{z}} = [\partial/\partial z_1, \partial/\partial z_2, \ldots, \partial/\partial z_N]^T$ for vector $\mathbf{z}$ and write for a function $g(\mathbf{z}, \mathbf{z}^*) : \mathbb{C}^N \times \mathbb{C}^N \to \mathbb{R}$, the first-order Taylor series expansion in terms of the two arguments of the function, z and z^* as

$$\triangle g = \langle \nabla_{\mathbf{z}^*} g, \triangle \mathbf{z}\rangle + \langle \nabla_{\mathbf{z}} g, \triangle \mathbf{z}^*\rangle = 2\mathrm{Re}\left\{\langle \nabla_{\mathbf{z}^*} g, \triangle \mathbf{z}\rangle\right\}. \quad (1)$$

Similarly, for the matrix gradient: $g(\mathbf{W}, \mathbf{W}^*) : \mathbb{C}^{N\times N} \times \mathbb{C}^{N\times N} \to \mathbb{R}$, we can write $\triangle g = 2\mathrm{Re}\{\langle \nabla_{\mathbf{W}^*} g, \triangle \mathbf{W}\rangle\}$ where $\nabla_{\mathbf{W}^*} g = \partial g/\partial \mathbf{W}^*$ is an $N \times N$ matrix whose (k, l)th entry is the partial derivative of g with respect to w_{kl}. It is also important to note that, in both cases, the gradient $\nabla_{\mathbf{z}^*} g$ defines the direction of the maximum rate of change in $g(\cdot, \cdot)$ with respect to $\mathbf{z}$, not $\nabla_{\mathbf{z}} g$, as sometimes incorrectly noted. It can be easily verified by using the Cauchy-Bunyakovski-Schwarz inequality that the term $\nabla_{\mathbf{z}^*} g$ leads to increments that are guaranteed to be nonpositive when minimizing a given function.

Relative Gradient:
We can use the matrix version of the expansion given in Eq. (1) to evaluate the relative [10] (or the natural [2]) gradient update rule, which has been usually simply extended to the complex case without proper justification [1], [4], [14]. To derive the relative gradient rule, a perturbation of $\mathbf{W}$ of the form $(\triangle\mathbf{W})\mathbf{W}$ is considered [10]. For the complex case, we can write the first-order Taylor series expansion as in Eq. (1) with the perturbation $(\triangle\mathbf{W})\mathbf{W}$ as

$$\begin{aligned}\triangle g &= \left\langle \frac{\partial g}{\partial \mathbf{W}^*}, (\triangle\mathbf{W})\mathbf{W}\right\rangle + \left\langle \frac{\partial g}{\partial \mathbf{W}}, (\triangle\mathbf{W}^*)\mathbf{W}^*\right\rangle \\ &= 2\mathrm{Re}\left\{\left\langle \frac{\partial g}{\partial \mathbf{W}^*}\mathbf{W}^H, \triangle\mathbf{W}\right\rangle\right\}\end{aligned}$$

and define the complex relative gradient of g at $\mathbf{W}$ as $(\partial g/\partial\mathbf{W}^*)\mathbf{W}^H$ to write the relative gradient update term as

$$\Delta\mathbf{W} = -\mu\frac{\partial g}{\partial\mathbf{W}^*}\mathbf{W}^H\mathbf{W}. \tag{2}$$

Upon substitution of $\Delta\mathbf{W}$ into (2), we observe that $\triangle g = -2\mu\cdot \|(\partial g/\partial\mathbf{W}^*)\mathbf{W}^H\|^2_{\mathrm{Fro}}$, *i.e.*, is a nonpositive quantity, thus a proper update term. In the next section, we show how the relative gradient update rule for complex ML ICA can be derived in a very straightforward manner using Eq. (2) and working in the framework we described in this section.

3. COMPLEX ICA BY MAXIMUM LIKELIHOOD AND MAXIMIZATION OF NONGAUSSIANITY

3.1. Complex Preliminaries

We first briefly introduce our notation and the relevant preliminaries for application to ICA. The joint probability density function (pdf) of a complex random variable $X = X_r + jX_i$ is defined as $p_X(x) \equiv p_{X_r X_i}(x_r, x_i)$ provided that it exists. Expectations of X are given by $E\{g(X)\} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x_r + jx_i)p_X(x)dx_r dx_i$ for any measurable function $g : \mathbb{C} \to \mathbb{C}$. We consider complex measurable functions, *i.e.*, functions for which the measure over the set of their singularities is zero in the complex vector field, and note that a random variable X is *circular in the strict-sense* if X and $Xe^{j\theta}$ have the same pdf.

The traditional ICA problem is considered such that $\mathbf{x} = \mathbf{A}\mathbf{s}$, $\mathbf{x}, \mathbf{s} \in \mathbb{C}^N$ and $\mathbf{A} \in \mathbb{C}^{N\times N}$. The task of the ICA algorithm is to determine a weight matrix $\mathbf{W}$ such that $\mathbf{u} = \mathbf{W}\mathbf{x} = \mathbf{P}\boldsymbol{\Lambda}\mathbf{s}$, where $\mathbf{P}$, a permutation matrix, represents the permutation ambiguity and $\boldsymbol{\Lambda}$, a diagonal matrix, represents the scaling ambiguity. Note that since the pdf of a complex random variable is defined through the joint density, to write the density of the observations $\mathbf{x}$ in terms of that of the source estimate, we need to consider the mapping $\mathbb{C}^N \mapsto \mathbb{R}^{2N}$ such that $\bar{\mathbf{u}} = \overline{\mathbf{W}}\bar{\mathbf{x}}$ where $\bar{\mathbf{x}} = [\mathbf{x}_r^T \mathbf{x}_i^T]^T$ and $\overline{\mathbf{W}} = \begin{bmatrix} \mathbf{W}_r & -\mathbf{W}_i \\ \mathbf{W}_i & \mathbf{W}_r \end{bmatrix}$ since $\mathbf{W} = \mathbf{W}_r + i\mathbf{W}_i$. The density of the transformed random variables is then written through the computation of the Jacobian as

$$p_X(\mathbf{x}) = |\det\overline{\mathbf{W}}| p_U(\mathbf{W}\mathbf{x}) \tag{3}$$

where $p_U(\mathbf{W}\mathbf{x}) = p_U(\mathbf{u}) = p_U(\mathbf{u}_r, \mathbf{u}_i)$.

ICA approaches that rely on nonlinear functions to implicitly generate the higher-order statistics to achieve independence offer practical and effective solutions to the ICA problem. In the next section, we consider ICA using maximum likelihood (ML) [19] (Infomax [8] and nonlinear decorrelations [1], [12] are closely related to the ML-based approach) and show how the framework described in Section 2 can be used to derive the relative (natural) gradient update rule for complex ICA using ML and introduce a number of score functions for adapting to the source distributions.

3.2. Complex Maximum Likelihood

Given T independent samples $\mathbf{x}(t) \in \mathbb{C}^N$, we can write the log-likelihood function as $\mathcal{L}(\mathbf{W}) = \sum_{t=1}^T \ell_t(\mathbf{W})$, where

$$\ell_t(\mathbf{W}) = \log p(\mathbf{x}(t)|\mathbf{W}) = \log p_S(\mathbf{W}\mathbf{x}) + \log|\det\overline{\mathbf{W}}|.$$

Here, we used the notation that $p_S(\mathbf{W}\mathbf{x}) \equiv \prod_{n=1}^N p_{S_n}(\mathbf{w}_n^H\mathbf{x})$, where $\mathbf{w}_n$ is the nth row of $\mathbf{W}$, $p_{S_n}(u_n) = p_{S_n}(u_{n_r}, u_{n_i})$ is the joint pdf of source n, $n = 1, \ldots, N$, and assumed that $\mathbf{W} = \mathbf{A}^{-1}$, *i.e.*, ignored the scaling and permutation ambiguity to write the likelihood directly in terms of $\mathbf{W}$ rather than the mixing matrix $\mathbf{A}$.

If a function $f(\cdot,\cdot)$ exists such that $p_S(\mathbf{u}_r, \mathbf{u}_i) = f(\mathbf{u}, \mathbf{u}^*)$ and satisfies the BAC, we can write

$$\frac{\partial \log f(\mathbf{u},\mathbf{u}^*)}{\partial\mathbf{W}^*} = \frac{\partial \log f(\mathbf{u},\mathbf{u}^*)}{\partial\mathbf{u}^*}\mathbf{x}^H \equiv -\psi(\mathbf{u},\mathbf{u}^*)\mathbf{x}^H \tag{4}$$

where $\mathbf{u} = \mathbf{W}\mathbf{x}$ and we have defined the score function $\psi(\mathbf{u},\mathbf{u}^*)$ that can be written directly by using Brandwood's theorem as

$$\psi(\mathbf{u},\mathbf{u}^*) = \frac{1}{2}\left(\frac{\partial \log p_S(\mathbf{u}_r,\mathbf{u}_i)}{\partial\mathbf{u}_r} + j\frac{\partial \log p_S(\mathbf{u}_r,\mathbf{u}_i)}{\partial\mathbf{u}_i}\right). \tag{5}$$

To compute $\partial\log|\det\overline{\mathbf{W}}|/\partial\mathbf{W}$, we first observe that $\partial\log|\det\overline{\mathbf{W}}| = \mathrm{Trace}(\overline{\mathbf{W}}^{-1}\partial\overline{\mathbf{W}}) = \mathrm{Trace}(\partial\overline{\mathbf{W}}\mathbf{P}\mathbf{P}^{-1}\overline{\mathbf{W}}^{-1})$, and then choose $\mathbf{P} = \frac{1}{2}\begin{bmatrix} \mathbf{I} & j\mathbf{I} \\ j\mathbf{I} & \mathbf{I} \end{bmatrix}$ to write

$$\begin{aligned}\partial\log|\det\overline{\mathbf{W}}| &= \mathrm{Trace}\left(\mathbf{W}^{-1}\partial\mathbf{W}\right) + \mathrm{Trace}\left((\mathbf{W}^*)^{-1}\partial\mathbf{W}^*\right) \\ &= \langle\mathbf{W}^{-H}, \partial\mathbf{W}\rangle + \langle\mathbf{W}^{-T}, \partial\mathbf{W}^*\rangle\end{aligned} \tag{6}$$

where we have used
$\mathbf{P}^{-1}\overline{\mathbf{W}}^{-1} = \frac{1}{2}\begin{bmatrix} \mathbf{W}^* & j\mathbf{W} \\ j\mathbf{W}^* & \mathbf{W} \end{bmatrix}^{-1} = \frac{1}{2}\begin{bmatrix} (\mathbf{W}^*)^{-1} & -j(\mathbf{W}^*)^{-1} \\ -j\mathbf{W}^{-1} & \mathbf{W}^{-1} \end{bmatrix}$.
We define $\Delta g(\mathbf{W},\mathbf{W}^*) \equiv \partial\log|\det\overline{\mathbf{W}}|$ and write the matrix version of the first-order Taylor series expansion Eq. (1) as

$$\begin{aligned}&\Delta g(\mathbf{W},\mathbf{W}^*) = \\ &\quad\left\langle\nabla_{\mathbf{W}^*}\log|\det\overline{\mathbf{W}}|, \Delta\mathbf{W}\right\rangle + \left\langle\nabla_{\mathbf{W}}\log|\det\overline{\mathbf{W}}|, \Delta\mathbf{W}^*\right\rangle,\end{aligned}$$

which, upon comparison with Eq. (6) gives us the required result for the matrix gradient:

$$\frac{\partial\log|\det\overline{\mathbf{W}}|}{\partial\mathbf{W}^*} = \mathbf{W}^{-H}. \tag{7}$$

We can then write the relative (natural) gradient updates to maximize the likelihood function using Eqs. (2), (4) and (7) as

$$\Delta\mathbf{W} = (\mathbf{W}^{-H} - \psi(\mathbf{u})\mathbf{x}^H)\mathbf{W}^H\mathbf{W} = (\mathbf{I} - \psi(\mathbf{u})\mathbf{u}^H)\mathbf{W}. \tag{8}$$

The update given above and the score function $\psi(\mathbf{u})$ defined in (5) coincides with those derived in [9] using a $\mathbb{C}^n \mapsto \mathbb{R}^{2n}$ isomorhic mapping in a relative gradient update framework and the one given in [14] considering separate derivatives. The derivation we have given here for the score function represents a very straightforward and simple evaluation compared to those in [9], [14], and more importantly shows how to bypass a major limitation in the development of ML theory for complex valued signal processing, that is working with probabilistic descriptions using complex algebra. Next, we demonstrate how the same framework can be used to introduce adaptive score functions and describe the properties of nonlinearities (score functions) previously employed for the update given in Eq. (8).

3.3. Score Functions

A true maximum likelihood scheme estimates both the parameters (in this case the demixing matrix $\mathbf{W}$) and the nonlinearity p_{S_n} to match the density of each source n. Typically in ML (or Infomax [8]) ICA, the form of the pdf (the nonlinearity) is fixed, or is chosen from two different nonlinearities depending on the sub- or super-Gaussian nature of the source estimate. A number of approaches for adapting the nonlinearity to the source estimate is proposed for the real-valued case. In [19], the score function is written as a linear combination of carefully selected basis functions, and these are extended to the complex case in [9] through complex-to-real mappings. In this section, we show how the formulation presented in this paper can be used to introduce score functions adapted to the source distributions by working completely in the complex domain.

The condition $p_S(\mathbf{u}_r, \mathbf{u}_i) = f(\mathbf{u}, \mathbf{u}^*)$ where $f(\mathbf{u}, \mathbf{u}^*)$ is analytic with respect to u and u^* independently, *i.e.*, that it satisfies the BAC is the key condition in the derivation given in Section 3.2. This condition translates into finding such a mapping for *each* source density $p_S(u)$ so that the condition for the joint multivariate density $p_S(\mathbf{u}_r, \mathbf{u}_i)$ is satisfied. Simple substitution of $u_r = (u + u^*)/2$ and $u_i = (u - u^*)/2j$ allows us to write a given pdf that is $\mathbb{R} \times \mathbb{R} \mapsto \mathbb{R}$ in terms of a function $\mathbb{C} \times \mathbb{C} \mapsto \mathbb{R}$ and almost all smooth functions that define a pdf can be shown to satisfy the BAC.

Score function based on the Gaussian pdf of order p:
The Gaussian density of order p introduced in [7] can be written as a function $\mathbb{C} \times \mathbb{C} \mapsto \mathbb{R}$ as such that

$$f_S(u, u^*) = \beta \exp(-(\alpha(u, u^*))^{p/2}/p(1 - \rho^2))$$

where $\alpha(u, u^*) = (u + u^*)^2/4\sigma_r^2 + j\rho(u^2 - u^{*2})/2\sigma_r\sigma_i - (u - u^*)^2/4\sigma_i^2$ with σ_r and σ_i are the standard deviations of the real and imaginary parts and ρ is the correlation coefficient between them. It satisfies the BAC, and for $p = 2$, $\beta = 1/2\pi\sigma_r\sigma_i\sqrt{1 - \rho^2}$, the pdf takes the form of the standard Gaussian and is super-Gaussian for $0 < p < 2$ and sub-Gaussian for $p > 2$. Note that when the sources are circular, *i.e.*, $\sigma_r = \sigma_i = \sigma$ and $\rho = 0$, $\alpha(u, u^*) = uu^*/\sigma^2$ and for the standard Gaussian we have the linear score function $\psi(u, u^*) = u/2\sigma^2$ as expected. The general score function for this pdf is simply evaluated as $\psi(u, u^*) = (1/2(1-\rho^2))(\partial\alpha/\partial u^*)\alpha^{(p-2)/2}$. In [17], the univariate form of this density is used to model the source densities for deriving ICA algorithms through negentropy maximization and significant performance gain is noted when the order p is updated during the estimation. Such a scheme can be adopted for ICA through ML as well and would also require the estimation of the variances of the real and imaginary parts of the sources.

Adaptive score function through linear combinations of bases:
In [9], the adaptive score functions of Pham and Garat [19] are extended to the complex case through $\mathbb{C}^N \mapsto \mathbb{R}^{2N}$ mappings. We can directly evaluate and write the adaptive scores in the complex domain as follows: Approximate the "true" score function $\psi_o(u, u^*)$ as a linear combination of M basis functions $\phi_m(u, u^*), m = 1, \ldots, M$ such that $\psi(u, u^*) = \sum_{m=1}^{M} \gamma_m^* \phi_m(u, u^*) = \boldsymbol{\gamma}^H \boldsymbol{\phi}$ where $\boldsymbol{\gamma} = [\gamma_1, \ldots, \gamma_M]^T$ and $\boldsymbol{\phi} = [\phi_1(u, u^*), \ldots, \phi_M(u, u^*)]^T$. Then, the problem is to determine the coefficient vector $\boldsymbol{\gamma}$ for each source such that $E\{|\psi_o(u, u^*) - \boldsymbol{\gamma}^H \boldsymbol{\phi}|^2\}$ is minimized. The solution is given by $\boldsymbol{\gamma} = (E\{\boldsymbol{\phi}\boldsymbol{\phi}^H\})^{-1} E\{\boldsymbol{\phi}\psi_o^*(u, u^*)\}$. The term $E\{\boldsymbol{\phi}\psi_o^*(u, u^*)\}$ requires that we know the true score function, which typically is not available. However, the useful observation is that when we write the expectation integral, substitute the expression for $\psi_o(u, u^*)$ given in (5) into the integral, we can evaluate this term as

$$E\{\boldsymbol{\phi}\psi_o^*(u, u^*)\} = 2E\left\{\frac{\partial \boldsymbol{\phi}}{\partial u^*}\right\} \tag{9}$$

provided that the function $g(u_r, u_i) = f_S(u, u^*)\phi(u, u^*)$ vanishes at infinity for u_r and u_i. Note that when evaluating the integral associated with the expected value computation, we consider the integral in the real-domain $\mathbb{R} \times \mathbb{R}$ for the subset of all values u and u^* such that u_r and u_i are real as in [3]. The factor 2 in Eq. (9) is due to the Jacobian of the transformation from the complex to the real domain.

In the real case, it is shown that if the set of basis functions contains at least the identity function plus some other non-linear function, then the stability of the separation is guaranteed [19]. A possible choice for the complex case is to use three functions such that $\boldsymbol{\phi} = [u, u^*, g(|u|)]^T$, which is similar to the set proposed in [9]. The adaptive score function formulation also provides a convenient way to incorporate prior information on the source distributions into the estimation.

Simple trigonometric and hyperbolic functions as the score function:
In [1], a number of complex trigonometric functions and their hyperbolic counterparts are introduced as the nonlinearity for achieving ICA in a "nonlinear decorrelations" framework [12], which is equivalent to the maximum likelihood updates when one nonlinearity is chosen as the identity and the second one as equivalent to the score function. Among those nonlinearities, especially the inverse tangent and sine—as well as their hyperbolic counterparts—are noted as good matches for super-Gaussian source distributions and functions such as $(-\text{asinh}(u) + u)$ as matches for sub-Gaussians. Again by defining integrals as in the evaluation of Eq. (9), we can thus determine the pdf implied by a chosen nonlinearity using the definition of the score function given in (4). Our numerical experiments based on these simple score functions as well as the two adaptive scores defined above demonstrated that these functions provide reliable and robust performance for a wide class of distributions, especially in the case of super-Gaussians.

4. DISCUSSION

As discussed in Section 3.3, score functions can be adapted to the source distributions through use of flexible bivariate pdfs such as the Gaussian densities of order-p or linear combinations of a number of basis functions. Such implementations are more effective when used in an adaptive framework that estimates the components one at a time and uses orthogonalization of the vectors $\mathbf{w}_n$ for subsequent estimations of the independent components as in [15], [17]. When all the components are estimated at the same time using an update of the form given in (8) fixing the nonlinearity and only adapting to the sub or super-Gaussian nature of the sources is a practical solution. The complex nonlinear score functions discussed in Section 3.3 using simple complex functions from the trigonometric and hyperbolic family present an attractive alternative for such a solution.

In Figure 1, the top figure shows the approximate density implied by the use of atanh u, *i.e.*, the magnitude of $\exp(-u^*\text{atanh}u)$. Even though the form of pdf that is implied is complex valued, which

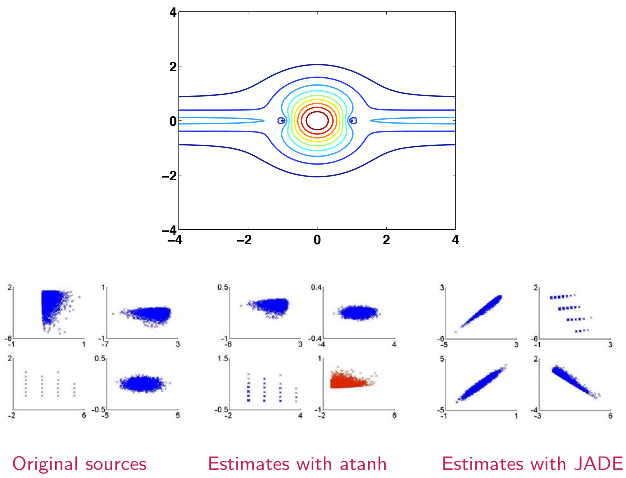

Fig. 1. *Top:* Form of pdf implied by the score function atanh; *Bottom:* Estimation results with ML using atanh score function and JADE for four noncircular sources.

is expected since the starting point has been the score function, the real part of the function dominates the overall response shown in the figure. The bottom figure shows the scatter plots for the estimation results of JADE [11] and ML with atanh as the score function. Note that the form of the pdf with atanh as the score has a dominant axis along the horizontal dimension, *i.e.*, the real part of the complex random variable. Thus, in the estimation results, we note that when the direction of the source matches with that of the pdf implied by the score function, the shape of the distribution of the estimated components is preserved, *i.e.*, the phase ambiguity that exists for the complex ICA is alleviated. Thus, even though the correlations of the magnitude with the original sources are close to unity for both JADE and ML-atanh in this example, the correlation of real and imaginary parts are high (close to unity) only for ML-atanh for the first three sources, those for which the direction of the source density matches with that of the nonlinearity. The red source estimate shows a rotated source estimate as its direction of asymmetry does not match with that of the density model given by atanh. These results were consistent over 100 different realizations of the source distributions.

We would like to thus note the richness of possible density matching mechanisms for the case of bivariate pdfs, *i.e.*, for the complex sources. When using other density matching mechanisms discussed in Section 3.3., there are a number of issues such as estimation of sources incrementally rather than all at once using an update as in (8) as well as whether to match the score function for all the sources or only for a few carefully selected sources identified using some prior information. Numerical results for these cases are not included here due to space constraints but as in the real case, we noted the advantage of using a few robust score functions instead of individually adapting to each source. The convergence speed is also significantly affected when adapting to the distributions of a large number of sources.

Maximum likelihood provides a desirable framework for many estimation problems, and in the case of ICA, it provides guidance for the choice of nonlinear functions to generate the higher-order statistics. In this paper, we show that we can establish a convenient framework for optimization in the complex domain using a result due to Brandwood [6] and can apply it to derive the relative gradient update rule for ICA using ML as well as for the selection of score functions working entirely in the complex domain. Such an approach provides an efficient framework and possibility to bypass assumptions commonly made in the derivation of complex ICA algorithms such as circularity [5], [20] and the use of real-to-complex transformations [9], [14], [21].

References

[1] T. Adalı, T. Kim, and V. D. Calhoun, "Independent component analysis by complex nonlinearities," in *Proc. ICASSP*, Montreal, Canada, 2004.

[2] S.-I. Amari, "Natural gradient works efficiently in learning," *Neural Computation*, vol. 10, pp. 251–276, 1998.

[3] P. O. Amblard, M. Gaeta, J. L. Lacoume, "Statistics for complex variables and signals – Part 1: Variables," *Signal Proc.*, vol. 53, pp. 1–13, 1996.

[4] J. Anemüller, T. J. Sejnowski, and S. Makeig, "Complex independent component analysis of frequency-domain electroencephalographic data," *Neural Networks*, vol. 16, pp. 1311–1323, 2003.

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

[6] D. Brandwood, "A complex gradient operator and its application in adaptive array theory," *Proc. IEE.*, vol. 130, pp. 11–16, Feb. 1983.

[7] S. De Simoni, "Su una estensione dello schema delle curve normali di ordine r alle variabili doppie," *Statistica*, vol. 37, pp. 447–474, 1968.

[8] A. J. Bell and T. J. Sejnowski, "An information maximisation approach to blind separation and blind deconvolution," *Neural Comput.*, vol. 7, pp. 1129–1159, 1995.

[9] J.-F. Cardoso and T. Adalı, "The maximum likelihood approach to complex ICA," in *Proc. ICASSP 2006*, Toulouse, France, 2006.

[10] J.-F. Cardoso and B. Laheld, "Equivariant adaptive source separation," *IEEE Trans. Signal Proc.*, vol. 44, pp. 3017–3030, Dec. 1996.

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

[12] A. Cichocki and R. Unbehauen, "Robust neural networks with online learning for blind identification and blind separation of sources" *IEEE Trans. Circuits Syst. I: Fund. Theory Apps.*, vol. 43, pp. 894–906, 1996.

[13] J. Eriksson and V. Koivunen, "Complex-valued ICA using second order statistics," in *Proc. MLSP*, Saõ Luis, Brazil, 2004.

[14] J. Eriksson, A. Seppola and V. Koivunen, "Complex ICA for circular and non-circular sources," in *Proc. EUSIPCO*, Antalya, Turkey, 2005.

[15] A. Hyvarinen, "Fast and robust fixed-point algorithms for independent component analysis," *IEEE Trans. Neural Networls*, vol. 10, pp. 626–634, 1999.

[16] T. Kim and T. Adalı, "Approximation by fully complex Multilayer Perceptrons" *Neural Comput.*, vol. 15, pp. 1641–1666, 2003.

[17] M. Novey and T. Adalı, "Adaptable nonlinearity for complex maximization of nongaussianity and a fixed-point algorithm," in *Proc. IEEE Workshop on Machine Learning for Signal Processing (MLSP)*, Maynooth, Ireland, Sep. 2006.

[18] 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, Sep. 2006.

[19] D.-T. Pham and P. Garat, "Blind Separation of Mixture of Independent Sources Through a Quasi-Maximum Likelihood Approach," *IEEE Trans. Signal Processing*, vol. 45, pp. 1712–1725, July 1997.

[20] P. Smaragdis, "Blind Separation of Convolved Mixtures in the Frequency Domain" *Neurocomputing*, vol. 22, pp. 21–34, 1998.

[21] A. van den Bos, "Complex gradient and hessian," *IEE Proc. Vision, Image, and Signal Processing,* vol. 141, no. 6, pp. 380–383, December 1994.