INDEPENDENT COMPONENT ANALYSIS BY COMPLEX NONLINEARITIES

Tülay Adalı[1], *Taehwan Kim,*[2] *and Vince Calhoun*[3,4]

[1] Dept. of CSEE, University of Maryland Baltimore County, Baltimore, MD, 21250
[2]The MITRE Corporation, McLean, VA, 22102
[3]Olin Neuropsychiatry Research Center, Institute of Living, Hartford, CT 06106
[4]Dept. of Psychiatry, Yale University, New Haven, CT 06520

ABSTRACT

A number of complex nonlinear functions are proposed for the independent component analysis (ICA) of complex-valued data. We discuss the properties of these nonlinearities and show their efficiency in generating the higher order statistics needed for ICA.

1. INTRODUCTION

Independent component analysis (ICA) for separating complex-valued sources is needed for convolutive source-separation in the frequency domain, or for performing source separation on complex-valued data, such as magnetic resonance imaging or radar data. For certain approaches to ICA, for example when using explicit tensor eigenvalue decomposition as in the ICA by joint approximate diagonalization of eigenmatrices (JADE) [6], the extension to the complex case is straightforward. However, in approaches that use nonlinear functions to implicitly generate the higher order statistics, such as InfoMax [3], this is not the case. These approaches provide simple and efficient solutions to the problem of ICA and thus it is desirable to extend them to process complex-valued data.

The main challenge for processing complex data with nonlinearities has been the conflict of boundedness and analyticity in the complex domain, as stated by Liouville's theorem. To overcome this challenge, there have been two major trends. The first one has been to define a complex 'split' nonlinear function such that the real and imaginary parts (or the magnitude and phase) of the argument are processed separately through real-valued nonlinear functions [11, 12]. Hence, boundedness is satisfied, but of course, at the expense of analyticity. The second approach processes the magnitude of the argument by a real-valued function [2, 4]. We argue, in this paper, that these approaches, while yielding satisfactory performance for a class of problems, are not effective in generating the higher order statistics required to establish independence when compared to complex nonlinear functions, *i.e.*, functions that are $\mathbb{C} \to \mathbb{C}$. We propose a number of elementary transcendental functions (ETFs) for ICA of complex-valued data and show that they are highly effective in extracting higher order statistical information. These functions are derivable from the entire exponential function e^z and hence are analytic. Also, with the use of complex nonlinear functions, a number of common simplifying assumptions, such as symmetricity of source distributions [2, 4] or uncorrelatedness of the real and imaginary parts of the sources [4] can be relaxed.

In [5], we derive a fully-complex counterpart of the Infomax algorithm by using a complex hyperbolic tangent function as the nonlinearity and note improved performance of the approach with respect to its split counterpart. While the argument in terms of generating higher order statistics to maximize the output of the entropy can be made for the fully-complex Infomax, the direct connection to maximum likelihood is no longer possible since the nonlinearity is complex-valued and hence cannot be directly associated with a source distribution. In this paper, we start directly with the update equation of Infomax that uses natural gradient, and by following the approach of [8], show the effectiveness of a class of complex nonlinearities in generating the higher order statistics required to achieve independence. Some of these nonlinearities, we note, are very robust, a somewhat surprising fact given their unbounded nature. We discuss the properties of these functions and give examples of their performance.

2. ICA BY COMPLEX NONLINEARITIES

2.1. Complex Preliminaries

A complex random variable X is defined as a random variable $X = \underline{X}_{re} + j\underline{X}_{im}$ where the real and imaginary parts, $\underline{X}_{re}$ and $\underline{X}_{im}$ are real-valued random variables. In our discussion, all real-valued variables and functions are underlined to distinguish them from complex variables and functions. The statistical properties of X are determined by the joint probability density function (pdf) $f_X(x) \equiv f_{\underline{X}_{re}\underline{X}_{im}}(\underline{x}_{re}, \underline{x}_{im})$, provided that it exists. The expectations are thus defined as

$$E\{g(X)\} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(\underline{x}_{re} + j\underline{x}_{im}) f_X(x) d\underline{x}_{re} d\underline{x}_{im}$$

for any measurable function $g : \mathbb{C} \to \mathbb{C}$. The ETFs we consider in this paper are all complex measurable functions similar to the case discussed in [9], *i.e.*, for these functions, the measure over the set of their singularities is zero in the complex vector field.

By a similar argument, we can extend the necessary and sufficient condition for the independence of two random variables to complex variables, *i.e.*, require that for two complex random variables X and Y,

$$E\{g(X)h(Y)\} = E\{g(X)\}E\{h(Y)\} \tag{1}$$

is satisfied for all measurable functions $g, h : \mathbb{C} \to \mathbb{C}$. Hence, two random variables X and Y are independent if and only if $g(X)$ and $h(Y)$ are uncorrelated for all functions g and h with finite expectations, *i.e.*, when equation (1) holds.

One more point to note for the subsequent discussion is that the uncorrelatedness of two complex random vectors $\mathbf{X}$ and $\mathbf{Y}$ requires that both their covariance and 'pseudo-covariance' matrices vanish, *i.e.*, $\mathbf{C}_{XY} = E\{(\mathbf{X} - E\{\mathbf{X}\})(\mathbf{Y} - E\{\mathbf{Y}\})^H\} = \mathbf{0}$ and $\tilde{\mathbf{C}}_{XY} = E\{(\mathbf{X} - E\{\mathbf{X}\})(\mathbf{Y} - E\{\mathbf{Y}\})^T\} = \mathbf{0}$, a necessary condition for the four possible combinations of the real and imaginary parts of X and Y, $E\{\underline{X}_{re}\underline{Y}_{re}\}$, $E\{\underline{X}_{im}\underline{Y}_{im}\}$, $E\{\underline{X}_{re}\underline{Y}_{im}\}$, and $E\{\underline{X}_{im}\underline{Y}_{re}\}$ to vanish.

2.2. ICA by Nonlinear Decorrelations

Consider the traditional ICA problem with linear mixing, *i.e.*, the vector of observed random variables $\mathbf{x} \in \mathbb{C}^n$ can be written as a linear mixture of the sources $\mathbf{s} \in \mathbb{C}^n$, such that $\mathbf{x} = \mathbf{A}\mathbf{s}$ where $\mathbf{A} \in \mathbb{C}^{n \times n}$. We assume that there are as many sources as mixtures and the data are preprocessed such that $E\{\mathbf{x}\} = 0$ and $E\{\mathbf{x}\mathbf{x}^H\} = \mathbf{I}$. The sources $\mathbf{s}$ are assumed to be statistically independent and the source estimate $\mathbf{u}$ is given by $\mathbf{u} = \mathbf{W}\mathbf{x}$, *i.e.*, $\mathbf{WA}$

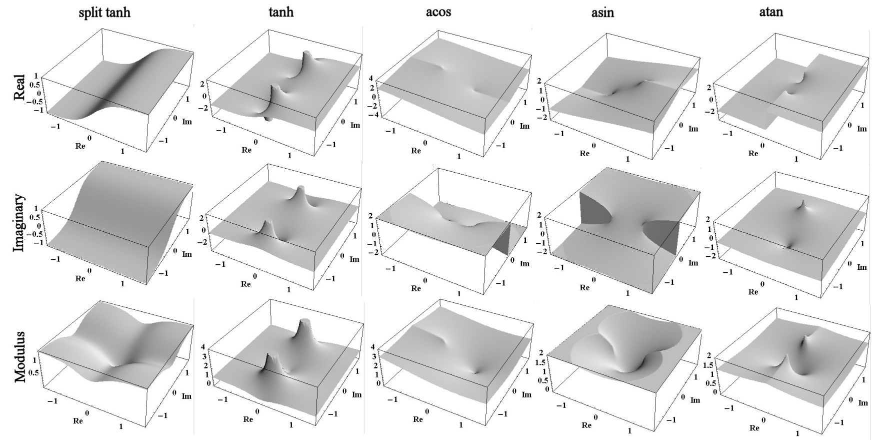

Fig. 1. Real, imaginary, and modulus characteristics of five complex functions

should approximate a permutation matrix. Note that for the complex case, an additional component of the scaling ambiguity is the phase of the sources since all variables are assumed to be complex-valued. Hence the permutation matrix will have one nonzero, unit magnitude element in each row.

In [5], we show that by using the complex $\tanh$ function, a complex-valued version of Infomax [3] can be derived and performs better than the approach using bounded but nonanalytic complex function defined by splitting the real and imaginary parts as $\tanh(\underline{u}_{re}) + j\tanh(\underline{u}_{im})$. The update for the complex Infomax using natural gradient is given by [5]

$$\Delta\mathbf{W} = \mu\left[\mathbf{I} - \varphi(\mathbf{u})\mathbf{u}^{H}\right]\mathbf{W} \tag{2}$$

where the nonlinearity is chosen as $\tanh(\mathbf{u})$ and in the update used as $\varphi(\mathbf{u}) = 2\tanh(\mathbf{u})$.

For the stationary solution of the learning rule in (2), *i.e.*, for $E\{\Delta\mathbf{W}\} = 0$,

$$E\{\varphi(u_k){u_l}^*\} = 0, \quad \text{for} \quad k \neq l \tag{3}$$

and

$$E\{\varphi(u_k){u_k}^*\} = 1 \tag{4}$$

where u_k is the kth element of $\mathbf{u}$. These conditions imply that the expression given in (1) is satisfied, as when the random vector $\mathbf{x}$ is centralized, $E\{\mathbf{u}\} = E\{\mathbf{Wx}\} = 0$. However, the condition in equation (1) should be satisfied for all $g(\cdot)$ and $h(\cdot)$. In the form given above, $h(\cdot)$ is the identity function. Hence, in order to approximate the condition in (1) by the conditions in (3) and (4), $\varphi(\cdot)$ should be rich enough to generate a wide range of higher order statistics. Assume that in a given region, the MacLaurin series expansion of $\varphi(u_k)$ converges and is written as $\sum_{i=0}^{\infty}\varphi_i u_k^i$ with φ_i as the coefficients for the ith power of u_k. The nonlinear decorrelation condition in (3) then implies $\sum_{i=0}^{\infty}\varphi_i E\{u_k^i u_l^*\} = 0$, *i.e.*,

$$E\{\left(\underline{u}_{re,k} + j\underline{u}_{im,k}\right)^i\left(\underline{u}_{re,l} - j\underline{u}_{im,l}\right)\} = 0$$

for $i = 0, 1, \ldots$, an expansion quite rich, generating a large number of decorrelation terms. This is needed in order to establish higher order decorrelations of different sources as well as of their real and imaginary parts as discussed in section 2.1. In contrast, the split type nonlinearity, e.g. as defined in [12], $\varphi(u_k) \equiv \underline{\varphi}(\underline{u}_{re,k}) + j\underline{\varphi}(\underline{u}_{im,k})$ implies

$$\sum_{i=0}^{\infty}\underline{\varphi}_i E\left\{(\underline{u}_{r,k}^i + j\underline{u}_{im,k}^i)(\underline{u}_{re,l} - j\underline{u}_{im,l})\right\} = 0,$$

a more limited expansion in producing the higher order terms.

In [8], ICA by using nonlinear decorrelations was proposed for real variables, with the update equation given as

$$\Delta\mathbf{W} = \mu\left[\mathbf{\Lambda} - g(\mathbf{u})(h(\mathbf{u}))^{H}\right]\mathbf{W}. \tag{5}$$

where $\mathbf{\Lambda}$ is a diagonal matrix. If $\mathbf{\Lambda}$ is chosen as the identity matrix, $g(\cdot)$ as $\varphi(\cdot)$, and $h(\cdot)$ as the identity function, the update becomes equal to the Infomax update with natural gradients. As shown by a number of authors, the Infomax approach is equivalent to maximum likelihood estimation when $\varphi(\cdot)$ matches the cumulative distribution function (cdf) of the sources. When complex functions are employed to generate the nonlinearities, direct connection to the cdf is lost as their cdf is defined as a real-valued function, the joint distribution of the real and imaginary parts. As we discuss in the following sections, there is a similar relationship between the characteristics of the nonlinear functions and the type of distributions they can approximate for the complex case as well, even though they can no longer be identified as cdfs.

2.3. Functions for Decorrelation

The following functions are noted to provide the nonlinear decorrelation required for ICA when used for the nonlinear function $\varphi(z)$ in the update given in (2).

Circular functions: $\tan z$ and $\cot z$,

Inverse circular functions: atan z, asin z, and acos z,

Hyperbolic functions: tanh z and coth z,

Inverse hyperbolic functions: atanh z, asinh z, and acosh z.

As expected, the trigonometric and the corresponding hyperbolic functions (tan z and tanh z *e.g.*) behave very similarly—their responses are simply $\pi/2$ rotated versions of each other. In particular, the functions: atan z, asin z, acos z, and tan z, and their hyperbolic counterparts performed consistently well over a wide range of input and mixtures, while the functions: sin z, cos z, acot z, sinh z, cosh z, and acoth z exhibited unstable behavior when used for ICA with the update given in (2) in our simulations.

In Fig. 1, we plot the characteristics of atan z, asin z, acos z, and tanh z along with split tanh $z \equiv \underline{\tanh}(\underline{z}_{re}) + j\underline{\tanh}(\underline{z}_{re})$, a nonanalytic function defined to satisfy the boundedness property in the complex domain [12]. As we show by examples in the simulations section, the inverse sine, cosine, and tangent—as well as their hyperbolic counterparts—provide particularly robust performance. Note that all these three functions, behave like an odd function, *i.e.*, they satisfy: acos $z = \pi -$ acos$(-z)$, asin $z = -$ asin$(-z)$, and atan $z = -$ atan$(-z)$ for all $z \in \mathbb{C}$. This behavior can be noted by the transitions from positive to negative in either the real or the imaginary response of these functions as shown in Fig. 1. Also, as an example, consider the MacLaurin expansion of inverse

Asymmetry (σ_r/σ_{im})	*Minimum Relative Entropy* split tanh	tanh	atanh	asin
1	0.105	0.280	0.067	0.059
3.3	0.309	0.121	0.095	0.026
33.3	0.949	0.110	0.081	0.026

Table 1. The minimum relative entropy distance of the output distribution to the uniform distribution for the nonlinearities shown in Fig. 2

tangent: atan$z = z - \frac{1}{3}z^3 + \frac{1}{5}z^5 - \frac{1}{7}z^7 \ldots$, given for all z, implying a structure rich in generating the higher order decorrelations. Also worthwhile to note is the squashing, or "cdf-like" characteristics of these functions in one of its components, the real or the imaginary, a property relevant to the discussions in the simulations section. Because of the sign reversal in the response of the inverse cosine function, it is used with a negative sign in the update equation (2), i.e., $\varphi(z) = -2\,\mathrm{acos}z$.

It is important to note that all functions that provided good performance–these shown in Fig. 1– are saturating type nonlinearities, or have decreasing rate of growth as we move away from the origin. This property, we conjecture, help these functions in reaching a stable point to satisfy equations (3) and (4). The inverse sine and cosine classes have branch cut-type singularities, while the tangent family possess point singularities that are periodic —at $(1/2+n)\pi, n \in$ N for tanh z and $(1/2+n)\pi j$ for tanh z— and the inverse tangent family isolated singular points—at $\pm j$ for atan z and at ± 1 for atanh z. All these singular points have measure zero and hence should not affect the performance in an implementation, an observation also verified in our simulations. The superior performance of acos z, asin z, and atan z suggests that the variability in their responses for the real and imaginary provides a richer structure in generating the higher order statistics to satisfy the independence of their outputs. Though their responses are not shown here, due to space limitations —a good resource for the properties of trigonometric and hyperbolic functions are the MathWorld pages at `http://mathworld.wolfram.com/`— it is worth noting that functions for which we observed unstable behavior such as the $\sin z$ and $\cos z$ have oscillatory type characteristics in their real and imaginary responses and have an increasing rate of growth as we move away from the origin.

Another way to assess the richness of a nonlinear function in generating higher order statistics is to look at the distribution of its output. As in the motivation for Infomax [3], the output of $\varphi(x)$ will approximate a uniform distribution if maximal information transfer is achieved through optimization of the weights $\mathbf{W}$. In Fig. 2, we plot the output distribution for four nonlinear functions, split $\tanh z$, $\tanh z$, atan z and asin z for a given single dimensional complex Gaussian input for three values of w, as well as the optimal output when the input distribution is highly asymmetric. The ability of the four functions to approximate a uniform distribution is quantified in Table 1. The minimum relative entropy distance of the output distribution from the uniform is given for the four nonlinear functions shown in Fig. 2. As observed in the figure and the table, all nonlinear functions perform well when the real and imaginary channels have the same variance, but the split-complex approach degrades significantly as the asymmetry increases. The three complex functions, especially the inverse hyperbolic tangent and the inverse sine provide superior performance in adapting to different input characteristics, which is also confirmed by our numerical simulation examples.

3. SIMULATIONS

The update given in (2) is tested with a number of complex and 'split-complex' type nonlinear functions from the trigonometric and hyperbolic family. For a wide range of source types and mixing matrices, most of the functions listed in section 2.3 yielded robust and satisfactory performance. We provide three example cases, particularly two for which performance degradation is observed.

$\frac{1}{2}\varphi(\cdot)$	*Correlation of modulus* *Source 1* kurt = 1.03	*Source 2* kurt = -0.57	*Source 3* kurt = 1.70
atan	0.99$\pm$0.02	0.98$\pm$0.05	1.0$\pm$0.0
atanh	0.99$\pm$0.02	0.98$\pm$0.05	1.0$\pm$ 0.0
asin	0.93$\pm$0.09	0.85$\pm$0.18	1.0$\pm$0.01
$-$acos	0.94$\pm$0.09	0.86$\pm$0.18	0.99$\pm$0.01
tanh	0.97$\pm$0.05	0.93$\pm$0.12	0.99$\pm$0.03
split tanh	1.0$\pm$0.0	1.0$\pm$0.0	1.0$\pm$0.0
split atan	1.0$\pm$0.0	1.0$\pm$0.0	1.0$\pm$0.0

Table 2. Correlation of the modulus of sources with the modulus of the estimates (all sources have symmetric distributions)

$\frac{1}{2}\varphi(\cdot)$	*Correlation of modulus* *Source 1* kurt= 4081	*Source 2* kurt= -0.72	*Source 3* kurt = 8.12×10^7
atan	0.99$\pm$0.05	0.91$\pm$0.11	1.0$\pm$0.0
atanh	0.99$\pm$0.0	0.93$\pm$0.0	1.0$\pm$0.0
asin	0.94$\pm$0.15	0.84$\pm$0.20	0.92$\pm$0.15
$-$acos	0.94$\pm$0.12	0.86$\pm$0.18	0.94$\pm$0.12
tanh$*$	0.45$\pm$0.28	0.21$\pm$0.17	0.81$\pm$0.09
split tanh	0.26$\pm$0.01	0.07$\pm$0.03	0.84$\pm$0.10
split atan	0.25$\pm$0.06	0.08$\pm$0.03	0.83$\pm$0.10

Table 3. Correlation of the modulus of sources with the modulus of the estimates ($*$ denotes that the learning rate is chosen smaller for convergence and the asymmetry of the distributions in terms of the ratio of the standard deviations σ_{re}/σ_{im} are given by 0.10, 0.35, and and 67.34.)

In the simulations, the source realizations are kept fixed and the complex-valued mixing matrix is varied for 50 independent runs. The sources are characterized in terms of their kurtosis values, the asymmetry of their real and imaginary distributions, and the correlation of their real and imaginary parts. The kurtosis is defined as in [4]: kurt$(s) = E\{|s|^4\} - 2\left(E\{|s|^2\}\right)^2 - \left|E\{s^2\}\right|$ for a zero mean random variable s. Hence, for a zero-mean unit variance Gaussian random variable with independent real and imaginary parts of equal variance, the kurtosis is zero. The mixtures are whitened to zero mean and unit variance prior to ICA, and an adaptive step size μ is used to help with convergence as in [3]. The correlation of the real and imaginary parts of the sources made little difference in performance, and in the three example cases we show the correlation of real and imaginary part of the sources is negligible. As would be expected, the performance of approaches assuming symmetric distributions as in [2] and [11] degraded when the condition for symmetric source distributions is not met.

For a large number of zero mean, unit variance sources with positive kurtosis most nonlinearities listed in Tables 2-4 provided robust and very good performance in terms of correlation with the modulus and the real and imaginary parts. Performance degradations are observed when sources with negative kurtosis are included in the mixture and when the sources are not normalized such that they have asymmetric standard deviations in real and imaginary. Examples for these cases are shown in Tables 2-4.

Also as the responses of the trigonometric functions and their hyperbolic counterparts are very similar —90 degrees rotated versions of each other— their performances in performing ICA were very similar as well. We show an example of this behavior for $\tanh$ and atanh in the tables. As noted in section 2.2, by using a complex nonlinearity, the direct connection to maximum likelihood estimation of the Infomax update is lost, however the performance differences for sources with negative and positive kurtosis suggest a similar extension is possible. For the case in Table 4, where all nonlinearities failed to provide a reasonable approximation to the sources with negative kurtosis, an 'extended' type nonlinearity that is matched to sub-Gaussian sources [10] (the form used in the simulations replaces the term in parentheses in (2) with $[1+\tanh(\mathbf{u})\mathbf{u}^H - \mathbf{u}\mathbf{u}^H]$) yielded better performance than the all

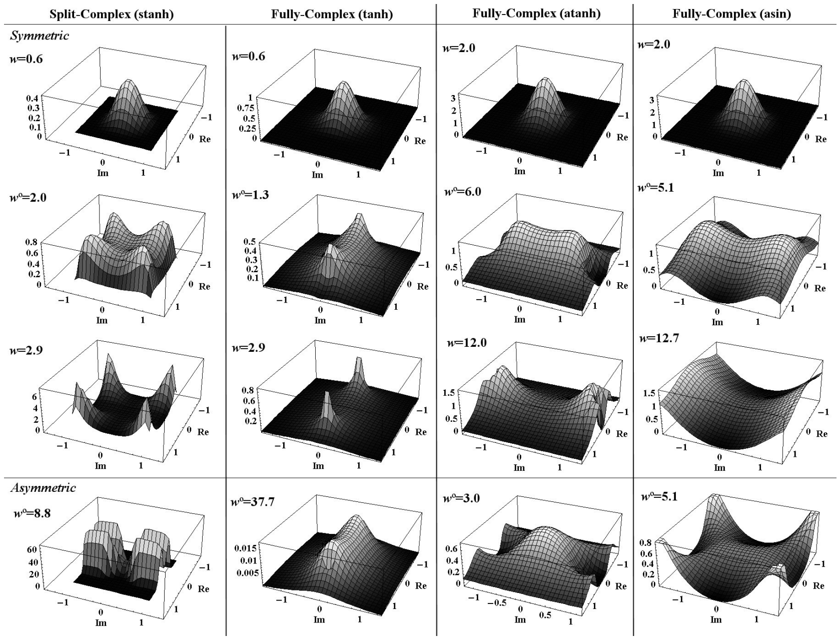

Fig. 2. Approximation by four different nonlinear functions

	Correlation of modulus		
	Source 1	*Source 2*	*Source 3*
$\frac{1}{2}\varphi(\cdot)$	kurt= -0.24	kurt= -0.51	kurt = 1.76
atan	0.64±0.18	0.62±0.19	1.0±0.0
atanh	0.60±0.13	0.56±0.13	1.0±0.0
asin	0.63±0.18	0.61±0.20	0.96±0.07
$-$acos	0.63±0.17	0.60±0.18	0.96±0.09
tanh	0.71±0.22	0.70±0.22	0.97±0.09
split tanh	0.45±0.04	0.41±0.04	1.0±0.0
split atan	0.47±0.06	0.43±0.06	1.0±0.0

Table 4. Correlation of the modulus of sources with the modulus of the estimates (all sources have symmetric distributions)

seven functions listed in Table 4, and the correlation of the modulus of the estimates for sources 1–3 for this case are: 0.86 ± 0.17, 0.89 ± 0.18, and 0.78 ± 0.16. Note that the performance of the extended nonlinearity is worse for the third source that has a positive kurtosis.

4. SUMMARY

We introduced a number of complex nonlinearities for ICA of complex valued sources. Because these functions are analytic, they satisfy the Cauchy-Riemann conditions, and their implementation is very straightforward. If they are used in a framework that starts from an objective function, their true gradient can be computed directly and is guaranteed to point in the right direction, direction of maximal change, as opposed to 'pseudo-gradients' of the split approach. In this paper, our starting point has been the update equation that generates higher-order statistics for independence of the estimates as in [8], an approach that yields exact same expression as the widely-used practical Infomax algorithm [3]. One of the reasons for taking this approach, starting by the update, has been the difficulty in providing a direct relationship to maximum likelihood for the complex nonlinearities, an argument that helps with the interpretation of nonlinearities as the cdf. The results we present however, note that a similar argument might be possible, though might not be direct as in the real-valued case. One possibility is to investigate an approach similar to that of [1] that uses *estimating functions*.

5. REFERENCES

[1] S-I. Amari and J.-F. Cardoso, "Blind source separation – semi-parametric statistical approach," *IEEE Trans. Acoust. Speech and Sig. Proc.*, vol. 45, no. 11, pp. 2692–2700, 1997.

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

[3] 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.

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

[5] V. Calhoun and T. Adalı, "Complex Infomax: Convergence and approximation of Infomax with complex nonlinearities," in *Proc. NNSP*, Martigny, Switzerland, Sep. 2002.

[6] J.-F. Cardoso, and A. Souloumiac, "Blind beamforming for non-Gaussian signals," *IEE Proc. Radar Signal Proc.*, vol. 140, no., 6, pp. 362–370, 1993.

[7] R.V. Churchill, J.W. Brown, and R.F. Verhey, Complex Variables and Applications, McGraw Hill, 1976.

[8] A. Cichocki and R. Unbehauen, "Robust neural networks with on-line learning for blind identification and blind separation of sources," *IEEE Trans. Circuits Syst.-I: Fund. Theory Apps.*, vol. 43, no. 11, pp. 894–906, 1996.

[9] T. Kim and T. Adalı, "Approximation by fully-complex multilayer perceptrons," *Neural Computation*, vol. 15, no. 7, pp. 1641–1666, 2003.

[10] T-W. Lee, M. Girolami, and T.J. Sejnowski, "Independent component analysis using an extended Infomax algorithm for mixed sub-Gaussian and super-Gaussian sources, *Neural Computation*, vol. 11, no. 2, pp. 417–441, 1999.

[11] H. Sawada, R. Mukai, S. Araki, and S. Makino, A polar-coordinate based activation function for frequency domain blind source separation," in *Proc. ICASSP*, Orlando, FL, May 2002, pp. I-1001–I-1004.

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