NONPARAMETRIC INDEPENDENT COMPONENT ANALYSIS FOR CIRCULAR COMPLEX VARIABLES

Intae Lee and Te-Won Lee

Institute for Neural Computation, University of California, San Diego
9500 Gilman Drive, La Jolla, CA 92093
{intelli, tewon}@ucsd.edu

ABSTRACT

A new consistent objective function for nonparametric complex independent component analysis (ICA) is proposed where the complex variables are restricted to be circular, or radially symmetric. This objective function is derived using an order statistics based density estimator which orders the complex data by their absolute values. The objective function is unconditional to the source distribution other than the circularity and it measures the statistical independence directly from the data. Using this objective function, a nonparametric complex independent component analysis algorithm can be derived. Also, the generalization of this objective function allows it to be combined with other algorithms to increase their separation performances. Experiments demonstrate the usefulness of the new objective function.

Index Terms— Array signal processing, multidimensional signal processing, frequency domain analysis

1. INTRODUCTION

Independent component analysis (ICA) is a well-known algorithmic method for separating statistically independent, or as independent as possible, source signals from their mixtures [1–3]. In the simplest form of the analysis, the model is an instantaneous mixture as

$$\mathbf{x}[n] = \mathbf{A}\mathbf{s}[n], \quad n = 1, 2, \cdots, N \tag{1}$$

where $\mathbf{x}[n]$ $(= [x_1[n], x_2[n], ...]^{\mathrm{T}})$, $\mathbf{s}[n]$ $(= [s_1[n], s_2[n], ...]^{\mathrm{T}})$, and $\mathbf{A}$ are the observation vector, the vector of mutually independent sources, and the mixing matrix, respectively. Here, we assume $\mathbf{A}$ to be square and invertible. The output we learn is denoted by $\mathbf{y}[n]$ $(= [y_1[n], y_2[n], ...]^{\mathrm{T}})$ and is obtained by

$$\mathbf{y}[n] = \mathbf{W}\mathbf{x}[n], \quad n = 1, 2, \cdots, N \tag{2}$$

where $\mathbf{W}$ is the estimate of $\mathbf{A}^{-1}$ and is called the unmixing matrix. While learning the unmixing matrix, most ICA algorithms regard each source signal $s_i[n], n = 1, 2, \cdots, N$ as N sampled data points of a random variable that follows certain probability distribution.

In the case when each source distribution is known and can be well represented by a function, e.g. Laplace distribution for speech signal, robust and simple ICA algorithms can be derived from maximum likelihood perspective. The prior information about the source distribution is brought in to the objective function, a.k.a. contrast, by nonlinearity functions where the corresponding source distribution is called source prior or source target. In many cases, however, it is certain that strong characterization of the source distributions is unavailable and an improper source target can result in poor separation result.

Instead of likelihood-equivalent contrasts, there are several ICA algorithms for real-valued variables that use more generalized contrasts. FastICA [4, 5], JADE [6], KernelICA [7], nonparametric ICA [8], order statistics based ICA [9], and RADICAL [10] are such algorithms. FastICA and JADE use some estimated measures of non-Gaussianity and fourth-order moments, respectively. KernelICA and nonparametric ICA can be regarded as tracking the unknown marginal densities of the data using kernel density estimation and updating the likelihood-like contrasts in each iteration by changing the source target accordingly. Order statistics based ICA and RADICAL use mutual information directly as their contrasts by using an order statistics based density estimator.

For circular complex variables, there are also several ICA algorithms that are blindly applicable to a wide range of source types, such as JADE and complex FastICA [11]. While JADE solves most 2×2 mixture problems, however, its performance degrades considerably when the number of mixed sources increases and also it is unable to separate sources with very small kurtoses. Complex FastICA is basically derived from a likelihood contrast with a fixed source target but nevertheless, it separates not only the source of the corresponding type but also a variety of sources of the other types. However, it occasionally finds some stationary points that are not the global optima.

Here, we propose a new nonparametric objective function for nonparametric complex ICA algorithms. This is an extension of order statistics based nonparametric ICA [9, 10] from real variables to complex variables. It is unconditional to the two-dimensional source distributions other than the circular dependency and has a variety of applications. With the given objective function, we derive a nonparametric complex ICA algorithm and show its usefulness and efficiency by simulation results.

2. ICA USING ORDER-STATISTICS BASED ENTROPY ESTIMATION

As mentioned earlier, there are generalized ICA algorithms that directly optimize a measure of the total statistical independence, the well-known mutual information,

$$\mathrm{I}(\mathbf{y}) = \mathrm{KL}(\mathrm{f}_{\mathbf{y}} || \prod_i \mathrm{f}_{y_i}), \tag{3}$$

where $\mathrm{f}_{\mathbf{y}}$ and f_{y_i} denote the joint density function and i-th marginal density of the output vector $\mathbf{y}$, respectively. This is done via order statistics based entropy estimation. From a given sample of N

This work was supported in part by National Science Foundation (NSF) grant IIS-0535251.

variates $x[1], x[2], \cdots, x[N]$ and its reordered set of $\bar{x}[n]$'s, $n = 1, 2, \cdots, N$ such that $\bar{x}[1] \leq \bar{x}[2] \leq \cdots \leq \bar{x}[N]$, $\bar{x}[j]$ is called the j-th order statistic. Using this, the probability density function (PDF) and the corresponding entropy of each random variable are estimated in simple forms [12, 13].

While D.-T. Pham first derived a nonparametric ICA algorithm using such entropy estimator with the order of spacing 1 [9], it was improved in RADICAL [12] to be more consistent and efficient by changing the spacing of order to m_N, a function of the data size N, and also by allowing overlaps of the spacing [10]. The contrast of RADICAL is as simple as

$$\min_{\mathbf{W}} \sum_i \sum_{n=1}^{N-m_N} \log(\bar{y}_i[n+m_N] - \bar{y}_i[n]) \tag{4}$$

where $\bar{y}_i[n]$'s denote the reordered set of the output data $y_i[n]$, $n = 1, 2, \cdots, N$ such that

$$\bar{y}_i[n] < \bar{y}_i[n+1], \quad \forall i, n = 1, 2, \cdots, N-1. \tag{5}$$

3. EXTENSION OF RADICAL FROM REAL VARIABLES TO COMPLEX VARIABLES

Density estimation using order statistics is not directly applicable to complex variables if the contour of its joint PDF in $\Re^2$ is unknown. In signal processing, we usually regard complex variables to live on circles, that is, we assume circularity. This assumption makes such density estimation for complex random variables feasible since then we can apply order statistics to complex-valued data by their absolute values, or the Euclidean distances from their center in $\Re^2$. This is depicted in Fig. 1.

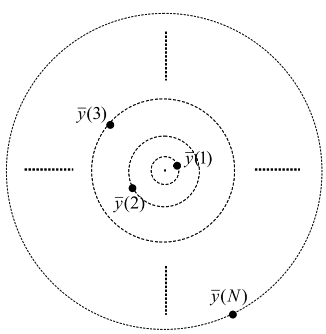

Fig. 1. How we apply order statistics to circular complex variables is depicted. The (zero-mean) complex data $y[1], y[2], \cdots, y[N]$ are ordered as $\bar{y}[1], \bar{y}[2], \cdots, \bar{y}[N]$ such that $\mathrm{abs}(\bar{y}[1]) < \mathrm{abs}(\bar{y}[2]) < \cdots < \mathrm{abs}(\bar{y}[N])$.

Since independence implies no correlation, many ICA algorithms keep the output data $\mathbf{y}[n]$ to be zero-mean and white ($\mathrm{E}[\mathbf{y}\mathbf{y}^H] = \mathbf{I}$) by prewhitening $\mathbf{x}[n]$ and constraining $\mathbf{W}$ to be orthogonal ($\mathbf{W}\mathbf{W}^H = \mathbf{I}$) mostly for fast learning and better performance (See [14]). For convenience, we will assume $\mathbf{y}[n]$ to be zero-mean and white from here on.

After introducing the differential entropy function of a random vector $\mathbf{v}$, which is defined as

$$\mathrm{H}(\mathbf{v}) = -\int_{\Re^{\dim(\mathbf{z})}} \mathrm{f}_{\mathbf{v}}(\mathbf{z}) \log \mathrm{f}_{\mathbf{v}}(\mathbf{z}) \mathrm{d}\mathbf{z}, \tag{6}$$

it can be shown that

$$\mathrm{I}(\mathbf{y}) = \sum_i \mathrm{H}(y_i) - \mathrm{H}(\mathbf{y}), \tag{7}$$

not only for real-valued y_i's but also for complex-valued y_i's (See [15]). Note that, in complex case, y_i in $\mathrm{H}(y_i)$ should be regarded as a real-valued two-dimensional vector ($\mathbf{y}_i$). Since $\mathrm{H}(\mathbf{y})$ on the right-hand-side of (7) is a constant value with respect to orthogonal $\mathbf{W}$, with the constraint of orthogonal $\mathbf{W}$ it holds that

$$\arg\min_{\mathbf{W}} \mathrm{I}(\mathbf{y}) = \arg\min_{\mathbf{W}} \sum_i \mathrm{H}(y_i). \tag{8}$$

The entropy term in (8) can be replaced with an entropy estimator. As previously discussed, we will apply order statistics to the complex-valued data by their absolute values in order to derive an entropy estimator for circular complex variables. Let's define $\bar{r}_i[n]$ and $\bar{y}_i[n]$ ($n = 1, 2, \cdots, N$) as the rearranged $|y_i[n]|$ and $y_i[n]$ ($n = 1, 2, \cdots, N$), respectively, such that

$$\bar{r}_i(n) < \bar{r}_i(n+1), \qquad n = 1, 2, \cdots, N-1, \tag{9}$$

$$\bar{r}_i(n) = |\bar{y}_i(n)|, \qquad n = 1, 2, \cdots, N. \tag{10}$$

Now we can follow the derivation of one-dimensional entropy estimator in [10] and derive an entropy estimator for circular complex random variables. For simpler notation, we will omit the subscript i while deriving the entropy estimator.

Let $\mathrm{F}_R(\cdot)$ be the cumulative distribution function that $\bar{r}$, or $|y|$, follows. Since $\mathrm{F}_R(\bar{r})$ is uniformly distributed in $[0, 1]$, it can be easily shown that

$$\mathrm{E}\left[\mathrm{F}_R(\bar{r}[n{+}1]) - \mathrm{F}_R(\bar{r}[n])\right] = \frac{1}{N{+}1}, \quad n = 1, \cdots, N{-}1. \tag{11}$$

Exploiting this idea, we approximate $\mathrm{f}_y(\cdot)$ as the following. We equally assign $\frac{1}{N+1}$ to each probability mass of the circular interval between two successive $\bar{r}(n)$'s and assume that each probability mass is uniformly distributed in the area that lies in each interval. Then

$$\hat{\mathrm{f}_y}(z) = \frac{\frac{1}{N+1}}{\pi(\bar{r}[n{+}1]^2 - \bar{r}[n]^2)}, \quad \bar{r}[n] \leq |z| < \bar{r}[n{+}1]. \tag{12}$$

An example of the PDF estimator is depicted in Fig. 2. In the figure we assumed four data points that correspond to $\bar{r}[j] = j$, $j = 1, 2, 3, 4$, and we arbitrarily added two end points of the support such that $\bar{r}[0] = 0$ and $\bar{r}[5] = 5$.

Now we can write

$$\begin{aligned}
&\mathrm{H}(y) \\
&= -\int_{\Re(z)} \int_{\Im(z)} \mathrm{f}_y(z) \log \mathrm{f}_y(z) \mathrm{d}\Im(z) \mathrm{d}\Re(z) && (13) \\
&= -\int_r 2\pi r \mathrm{f}_y(z) \log \mathrm{f}_y(z) \mathrm{d}r && (14) \\
&= -2\pi \sum_{n=0}^{N} \int_{\bar{r}[n]}^{\bar{r}[n+1]} r \mathrm{f}_y(z) \log \mathrm{f}_y(z) \mathrm{d}r && (15) \\
&\approx -2\pi \sum_{n=1}^{N-1} \int_{\bar{r}[n]}^{\bar{r}[n+1]} \frac{\frac{r}{\pi(N+1)}}{\bar{r}[n{+}1]^2 - \bar{r}[n]^2} \log\left(\frac{\frac{1}{\pi(N+1)}}{\bar{r}[n{+}1]^2 - \bar{r}[n]^2}\right) \mathrm{d}r && (16) \\
&= c_a \sum_{n=1}^{N-1} \log(\bar{r}[n{+}1]^2 - \bar{r}[n]^2) + c_b && (17)
\end{aligned}$$

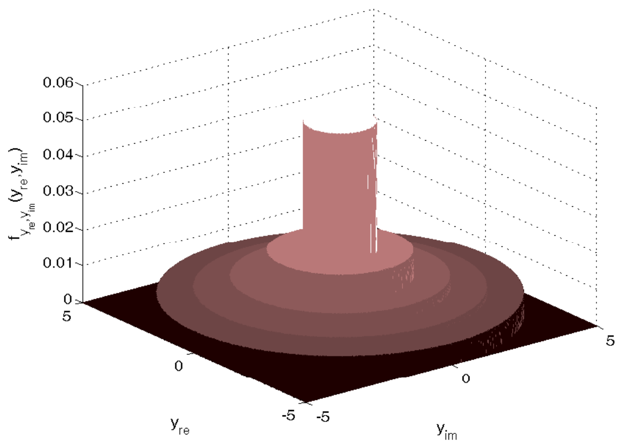

Fig. 2. A PDF estimated by the density estimator for circular complex random variables (12) where $\bar{r}[i] = i$, $i = 0, 1, \cdots, 5$, and $\bar{r}(0)$ and $\bar{r}(5)$ correspond to the end points of the support.

where $\Re(z)$ and $\Im(z)$ are respectively the real and imaginary parts of the complex variable z. Also $r = |z|$, and c_a and c_b are (and will be) used to denote some constants whose values are not important. The approximation in (16) comes from replacing $\mathrm{f}_y(y)$ with $\hat{\mathrm{f}}_y(y)$ in (12) and from removing the sum for n equaling 0 and N since the end points of $\mathrm{f}_y(y)$'s support, $\bar{r}(0)$ and $\bar{r}(N)$, are usually unknown.

In order to smoothen and reduce the variance of the estimator, two additional steps will be taken. First, we change the order of spacing from 1 to m_N, which results in

$$\hat{\mathrm{H}}(y;j) = c_a \sum_{n=1} \log(\bar{r}[m_N n+j]^2 - \bar{r}[m_N(n-1)+j]^2) + c_b \qquad j = 1, 2, \cdots, m_N. \quad (18)$$

Note that there are m_N different entropy estimators with respect to the choice of the first datum of the spacing. Second, we take the arithmetic mean of all of the shifted entropy estimators in (18) to obtain the following simple but consistent entropy estimator

$$\hat{\mathrm{H}}(y) = \frac{1}{m_N} \sum_{j=1}^{m_N} \hat{\mathrm{H}}(y;j) \quad (19)$$

$$= c_a \sum_{n=1}^{N-m_N} \log(\bar{r}[n+m_N]^2 - \bar{r}[n]^2) + c_b. \quad (20)$$

Plugging the entropy estimator in (20) into the entropy term on the right-hand-side of (8), we obtain the following objective function with the constraint of $\mathbf{W}$ being orthogonal;

$$\mathbf{W}_{\mathrm{opt}} = \arg\min_{\mathbf{W}} \sum_i \sum_{n=1}^{N-m_N} \log(\bar{r}_i[n+m_N]^2 - \bar{r}_i[n]^2). \quad (21)$$

As a simple extension of RADICAL, this contrast inherits the advantages, i.e. consistency and efficiency. For this, the following conditions should hold (See [10, 12, 13]);

$$m_N, N \to \infty, \quad \frac{m_N}{N} \to 0. \quad (22)$$

It is typical for m_N to be $\sqrt{N}$.

4. APPLICATIONS AND EXPERIMENTS

Now, we can use the new contrast in (21) to separate independent circular complex signals from their mixtures. Note that instantaneous mixtures of independent circular complex variables are also circular and hence our measure of mutual independence is still applicable to their mixtures. By following the notations in [16] and defining $\mathbf{w}_i$ to be the i-th row of $\mathbf{W}$, we derive the following gradient descent update rule;

$$\triangle \mathbf{w}_i \propto -\frac{\partial}{\partial(\mathbf{w}_i)^*} \sum_i \sum_{n=1}^{N-m_N} \log(\bar{r}_i[n+m_N]^2 - \bar{r}_i[n]^2) \quad (23)$$

$$= \sum_{n=1}^{N-m_N} \frac{\bar{y}_i[n+m_N](\bar{\mathbf{x}}_i[n+m_N])^{\mathrm{H}} - \bar{y}_i[n](\bar{\mathbf{x}}_i[n])^{\mathrm{H}}}{\bar{r}_i[n+m_N]^2 - \bar{r}_i[n]^2}. \quad (24)$$

where $\bar{\mathbf{x}}_i[n]$'s are the ordered $\mathbf{x}[n]$'s such that $\mathbf{w}_i \bar{\mathbf{x}}_i[n] = \bar{y}_i[n]$.

Because our contrast is based on order statistics and thus is a function of ordered data where the order easily changes after each update, it has a large number of potential local optima. Hence the gradient descent update rule does not guarantee convergence on the global optimum. While RADICAL employs an exhaustive rotational search, here, for our algorithm, we propose the following exhaustive search that uses the gradient descent update in (24).

1. Starting from a big learning rate, $\mathbf{W}$ is being updated by gradient descent and the $\mathbf{W}$ that results in the smallest value for the contrast is kept track of.
2. The learning rate is reduced and the update of $\mathbf{W}$ starts over from the one that resulted in the smallest contrast value in the previous step.
3. 2) is repeated.

Note that, $\mathbf{W}$ is constrained to be orthogonal and the data set is resorted after every update. The performance of the nonparametric complex ICA algorithm is tested with synthetic data. For each independent source signal, we chose 10000 complex-valued data points that are uniformly distributed along circles in $\Re^2$. Simple 2×2 problems were generated by mixing those source signals with a randomly generated mixing matrix A. Here, we will show examples when our algorithm outperforms complex JADE and complex FastICA.

First, while complex JADE failed in identifying circular complex signals with small kurtoses from their mixtures, the nonparametric complex ICA algorithm separated them very well. Second, we could find cases where complex FastICA frequently found the wrong stationary points and the new algorithm did not fail. This is shown in Fig. 3 by scatter plots. The scatter plots of the arbitrarily generated independent source data, the mixtures of the source data, the wrong separation result of complex FastICA, and the separation result of the nonparametric complex ICA algorithm are respectively shown in (a), (b), (c), and (d). While complex FastICA resulted in wrong output signals as in Fig. 3(c) 33 times out of 100 trials, nonparametric ICA found the sources correctly as in Fig. 3 (d) all 100 times.

Although the nonparametric complex ICA algorithm shows excellent separation results for circular complex signals, the search is exhaustive and also it is still possible that the algorithm misses the correct answer during its search. Another way to use the new objective function is to compare the results of one or various algorithms with respect to the values of the nonparametric independence measure and then choose the best answer. For example, the new objective function can complement the performance of complex FastICA which sometimes finds the wrong optima as in our experiment.

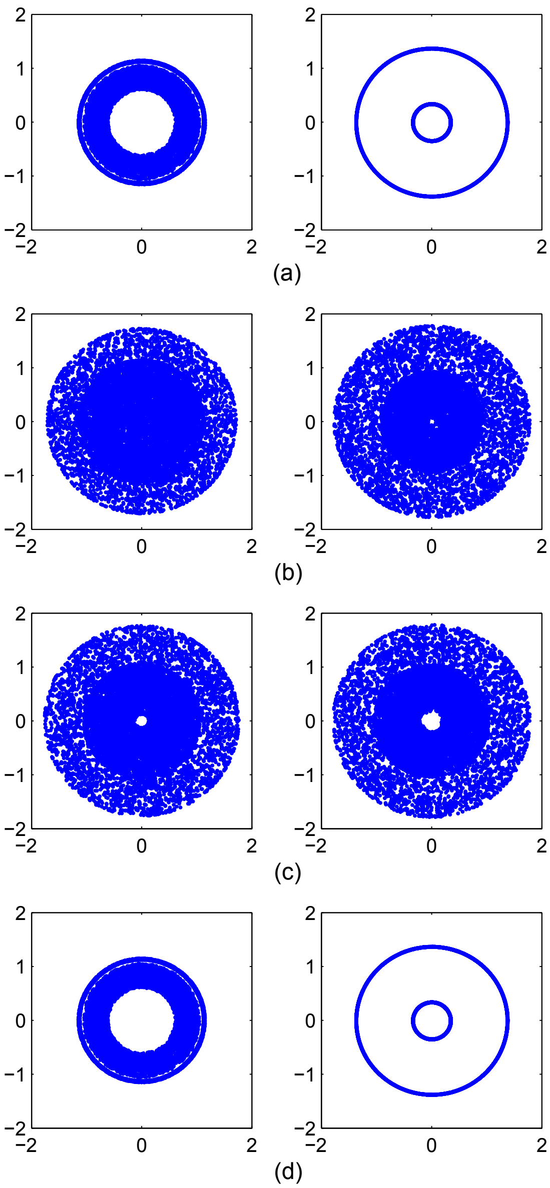

Fig. 3. The separation results of complex FastICA and nonparametric complex ICA are compared. The scatter plots of the arbitrarily generated independent source data, the mixtures of the source data, the typical wrong separation result (33 times out of 100 trials) of complex FastICA, and the separation result of the nonparametric complex ICA algorithm are respectively shown in (a), (b), (c), and (d)

5. DISCUSSION

For the derivation of the new objective function, we assumed circular symmetry for the complex variables. Since the complex-valued frequency components of a time-domain signal show circularity, the objective function is applicable to any time-domain signal in the frequency domain. Also, the objective function can easily be extended to even higher dimension where the multidimensional variables have the property of spherical invariance, e.g. the frequency components of some natural signals including speech. Such an application can be seen in [17].

6. REFERENCES

[1] J. Herault and C. Jutten. Space or time processing by neural network models. In *Proc. AIP Conf.: Neural Networks for Computing*, volume 151, 1986.

[2] P. Comon. Independent component analysis-a new concept? *Signal Processing*, 36:287–314, 1994.

[3] A. J. Bell and T. J. Sejnowski. An information maximization approach to blind separation and blind deconvolution. *Neural Computation*, 7(6):1129–1159, 1995.

[4] A. Hyvärinen and E. Oja. A fast fixed-point algorithm for independent component analysis. *Neural Computation*, 9(7):1483–1492, 1997.

[5] A. Hyvärinen. Fast and robust fixed-point algorithms for independent componenet analysis. *IEEE Trans. on Neural Networks*, 10(3):626–634, 1999.

[6] J.-F. Cardoso. High-order contrasts for independent component analysis. *Neural Computation*, 11(1):157–192, 1999.

[7] F. R. Bach and M. I. Jordan. Kernel independent component analysis. *J. of Machine Learning Research*, 3:1–48, 2002.

[8] R. Boscolo, H. Pan, and V. P. Roychowdhury. Independent component analysis based on nonparametric density estimation. *IEEE Trans. on Neural Networks*, 15(1):55–65, 2004.

[9] D.-T. Pham. Blind separation of instantaneous mixture of sources based on order statistics. *IEEE Trans. on Signal Processing*, 48(2):363–375, 2000.

[10] E. G. Learned-Miller and J. W. Fisher III. ICA using spacings estimates of entropy. *J. of Machine Learning Research*, 4:1271–1295, 2003.

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

[12] O. Vasicek. A test for normality based on sample entropy. *J. of the Royal Statistical Society, Series B*, 38(1):54–59, 1976.

[13] J. Beirlant, E. J. Dudewicz, L. Györfi, and E. C. van der Meulen. Nonparametric entropy estimation: An overview. *Int. J. of Mathematical and Statistical Sciences*, 6:17–39, 1997.

[14] A. Hyvärinen and E. Oja. *Independent Component Analysis*. John Wiley and Sons, 2002.

[15] I. Lee, T. Kim, and T.-W. Lee. Fast fixed-point independent vector analysis algorithms for convolutive blind source separation. *To appear in Signal Processing*, 2007.

[16] A. van den Bos. Complex gradient and hessian. In *IEE Proceedings on Vision, Image and Signal Processing*, volume 141, pages 380–382, 1994.

[17] I. Lee and T.-W. Lee. On the assumption of spherical symmetry and sparseness for the frequency-domain speech model. *To appear in IEEE Trans. on Speech, Audio and Language Processing*, 2007.