AN EFFICIENT ENTROPY RATE ESTIMATOR FOR COMPLEX-VALUED SIGNAL PROCESSING: APPLICATION TO ICA

Geng-Shen Fu, Ronald Phlypo, Matthew Anderson, Xi-Lin Li, and Tilitay Adalı

University of Maryland, Baltimore County, Dept. of CSEE, Baltimore, MD 21250

ABSTRACT

Estimating likelihood or entropy rate is one of the key issues in many signal processing problems. Mutual information rate, which leads to the minimization of entropy rate, provides a natural cost for achieving blind source separation (BSS). In many complex-valued BSS applications, the latent sources are non-Gaussian, noncircular, and possess sample dependence. Consequently, an effective estimator of entropy rate that jointly considers all three properties of the sources is required. In this paper, we propose such an entropy rate estimator that assumes the sources are generated by invertible filters. With this new entropy rate estimator, we propose a complex entropy rate bound minimization algorithm. Simulation results show that the new method exploits all three properties effectively.

Index Terms— Independent component analysis, Mutual information rate, Entropy rate.

1. INTRODUCTION

Independent component analysis (ICA) has been one of the most attractive solutions for the BSS problem because of its widely applicable assumption on independence. ICA can estimate a demixing matrix and separate signals under the assumption of statistical independence among the source signals. Furthermore, complex-valued ICA (CICA) is widely used in a number of applications such as communications, radar, and biomedicine [1–3].

In general, a CICA can be achieved by exploiting the following three types of diversity—signal property: non-Gaussianity (higher-order statistics), sample dependency, or noncircularity [2, 4–6]. But, to the best of our knowledge, all of the existing CICA algorithms exploit only one or two of the diversities, and most of them ignore sample dependence. The strongly uncorrelating transform (SUT) [7, 8] and the generalized uncorrelating transform (GUT) [9] algorithms only make use of noncircularity by only using second-order statistics. The joint approximate diagonalization of eigenmatrices (JADE) [10], complex FastICA [11], and algorithms using nonlinear functions [12, 13] only exploit non-Gaussianity by assuming source is circular explicitly or implicitly. Non-circular FastICA [14], the complex fixed-point algorithm (CFPA) [15], kurtosis maximization (KM) [16], entropy bound minimization (CEBM) [17], and algorithms using nonlinear functions [18–21] take both non-Gaussianity and noncircularity into account. The Gaussian entropy rate minimization algorithm [22] separates noncircular correlated sources by exploiting non-Gaussianity and sample dependence.

In this paper, we present a new algorithm that takes all three types of diversity into account. By assuming each source is generated by an invertible filter driven by an independently and identically distributed (i.i.d.) random process, we propose an entropy rate estimator. Then, we use the mutual information rate to derive an algorithm, CICA by entropy rate bound minimization (CERBM), that exploits all three types of diversity. Instead of minimizing entropy rate, which is equivalent to the minimization of mutual information rate, we minimize a bound on the entropy rate using a semi-parametric method. By estimating an upper bound, we achieve an algorithm which is robust to model mismatch. We show that CERBM makes use of all three types of diversity and provides very desirable performance by comparing its performance with those of competing algorithms.

2. BACKGROUND

2.1. Preliminaries

We assume that a complex random vector \( \mathbf{x} = \mathbf{x}_R + j\mathbf{x}_I \in \mathbb{C}^T \) has zero mean, where \( j = \sqrt{-1} \) is the imaginary unit. Let \( \mathbf{x} = [\mathbf{x}^T, \mathbf{x}_I^H]^T \) be the complex augmented vector, where superscript \( T \) and \( H \) denote the transpose and Hermitian, respectively. The second-order statistics are given by the following augmented covariance matrix

\[
\mathbf{R} \triangleq E \{ \mathbf{x} \mathbf{x}^H \} = \begin{bmatrix} \mathbf{R} & \overline{\mathbf{R}} \\ \mathbf{R}^* & \overline{\mathbf{R}}^* \end{bmatrix},
\]

where \( E \) denotes expectation, superscript * denotes complex conjugate, and \( \mathbf{R} \triangleq E \{ \mathbf{x} \mathbf{x}^H \} \) and \( \overline{\mathbf{R}} \triangleq E \{ \mathbf{x} \mathbf{x}^T \} \)
are called the covariance and pseudo-covariance (or complementary covariance) matrix, respectively.

The probability density function (PDF) and entropy rate of \( x \) are defined as \( p(x) \equiv p(x|y, x_i) \) and \( H_r(x) \equiv H_r(x_i, x) = \lim_{T \to \infty} E \{-\log p(x_i, x_T)\} / T \), respectively. The random vector \( x \) is called second-order circular or proper if pseudo-covariance matrix \( \hat{R} = 0 \) and strictly circular, or circular, if \( x \) and \( x \in \Theta \) have the same PDF for any \( \Theta \in \mathbb{R} [1] \).

### 2.2. ICA model and cost function

Let \( N \) statistically independent zero-mean sources \( s(t) = [s_1(t), \ldots, s_N(t)] \) be mixed through an \( N \times N \) nonsingular mixing matrix \( A \in \mathbb{C}^{N \times N} \) so that we obtain the mixtures \( x(t) = [x_1(t), \ldots, x_N(t)]^T \).

\[
x(t) = As(t), 1 \leq t \leq T,
\]

where \( t \) is the discrete sample index. The mixtures are separated as \( y(t) = Wx(t) \), where \( W = [w_1, \ldots, w_N] \) is the demixing matrix, which is the quantity to be estimated for ICA, and \( y(t) \equiv [y_1(t), \ldots, y_N(t)]^T \). This can also be written in matrix form, \( Y = WX = W S \), where \( S = [s_1, \ldots, s_N]^T \in \mathbb{C}^{N \times T} \) and \( s_i \in \mathbb{C}^T \). The same definition holds for \( x \) and \( y \).

A natural cost for achieving the separation of these independent sources is the mutual information rate \( \mathcal{I}_r = \sum_{i=1}^{N} H_r(y_i) - \log \det(WW^H) - H_r(x) \) among random processes \( y_i \), \( i = 1, \ldots, N \), where \( H_r(y_i) \) is the entropy rate of the \( i \)th process \( y_i \), and the entropy rate of the vector process \( x \), \( H_r(x) \), is a constant with respect to \( W \). Hence, the cost function is given by:

\[
\mathcal{J}_r(y_1, \ldots, y_N) = \sum_{i=1}^{N} H_r(y_i) - 2 \log |\det(W)|, \tag{1}
\]

since \( \det(WW^H) = |\det(W)|^2 \). Mutual information rate cost includes all three types of diversity, since entropy rate is defined by the joint PDF of the whole complex random process. If the samples are i.i.d., this cost function will reduce to mutual information \( I(y_1; \ldots; y_N) \), and, as noted in [4, 18, 23], the mutual information rate cost function is intimately related with the maximum likelihood and non-Gaussianity cost.

### 3. COMPLEX ENTROPY RATE ESTIMATOR

For the estimation of entropy rate, we assume that there exists a whitening filter \( \hat{b} \) such that the output process \( z \) will then be an i.i.d. process. We can always scale the whitening filter \( \hat{b} \) such that the input and output will have equal entropy, which means the entropy rate of \( y \) equals the entropy of \( z \) since \( z \) is i.i.d. The optimum filter coefficients \( \hat{b} \) can be obtained by solving the following optimization problem:

\[
\min_b H(z), \text{s.t. } |p_0|^2 - |q_0|^2 = 1, \tag{2}
\]

where the constraint \( |p_0|^2 = |q_0|^2 = 1 \) makes sure that the input and output will have equal entropy [22]. Estimation of the entropy of a complex random variable requires estimation of a bivariate distribution which is more complicated than the univariate real-valued case. Thus, we estimate an entropy bound \( H(z) \leq H(z_R) + H(z_I) \) instead, where the equality holds if and only if the real and imaginary part of \( z \) are independently distributed.

For the estimation of entropy of a real-valued random variable, we use a suite of maximum entropy distributions to form a flexible model and can approximate the entropies of a wide range of distributions, including sub-Gaussian, super-Gaussian, unimodal, bimodal, symmetric and skewed distributions [24]. The entropy bound of a real-valued random variable \( n \) with unit variance is given by \( H(n) = 0.5 \log(2\pi e) - V(E\{G(n)\}) \), where \( G(\cdot) \) is a measuring function for a maximum entropy distribution, and \( V(\cdot) \) is the negentropy defined in [24].

Hence, the cost function \( H(z) \) in (2) can be derived using

\[
H(z) \leq \log(\sigma_{z_R}\sigma_{z_I}) + H(\tilde{z}_R) + H(\tilde{z}_I), \tag{3}
\]

where \( \tilde{z}_R = z_R/\sigma_{z_R}, \tilde{z}_I = z_I/\sigma_{z_I}, \) and \( \sigma_{z_R} \) and \( \sigma_{z_I} \) are standard deviation of \( z_R \) and \( z_I \), respectively. The problem can then be written as the following Lagrangian function

\[
L_b(b, \lambda) = \log(2\pi e\sigma_{z_R}\sigma_{z_I}) - V_1(E\{G_1(z_R)\})
- V_2(E\{G_2(\tilde{z}_I)\}) + \lambda \left( b^H \mathbf{D}_b b - 1 \right), \tag{4}
\]

where

\[
\mathbf{D}_b = \text{diag} \left( \begin{array}{cccc}
1 & 0 & \cdots & 0
0 & -1 & \cdots & 0
\vdots & \vdots & \ddots & \vdots
0 & 0 & \cdots & -1
\end{array} \right)_{K \times K}.
\]

For updating, we use Wirtinger calculus [1, 25] to compute the gradient of the Lagrangian function (4) with respect to \( b \). The details of the derivations are given in Appendix A. Hence, the entropy rate of \( y, H_r(y) \), is bounded by \( H(z_R) + H(z_I) \).

In [17], the bound \( H(u) + H(v) \) is used for \( H(z) \), where uncorrelated random variables \( u \) and \( v \) are linearly transformed from \( z_R \) and \( z_I \). In our case, the whitening filter \( b \) makes the samples of \( z \) to be i.i.d., and also makes \( z_R \) and \( z_I \) as independent as possible, since we use \( H(z_R) + H(z_I) \) as our cost function. Also, it can
be shown that a widely linear filter followed by a linear transformation is always equivalent to another widely linear filter. Hence, theoretically, we may achieve a tighter entropy bound by using the widely linear whitening filter output $z_R$ and $z_I$, rather than $u$ and $v$.

4. CERBM ALGORITHM

Using mutual information rate cost function (1), we propose an algorithm that exploits all three types of diversity: non-Gaussianity, sample dependency, and non-circularity. Using the new entropy rate estimator (3), the entropy rate $H_r(y_t)$ is estimated by $H(z_i)$. Filter coefficients for different whitening filters can be different, but filter orders are assumed to be the same for simplicity.

Instead of minimizing $J_r(W)$ with respect to the demixing matrix $W$, we use a decoupling procedure [26, 27] to divide the problem into minimizing $J_r(W)$ with respect to each of the row vectors $w_i$, $i = 1, \ldots, N$. We can then write the cost as a function of only $w_i$, which is

$$J_i(w_i) = H(z_i) - 2 \log |h_i^H w_i| + C_i$$

$$\leq \log(2\pi e) + \log \sigma_{z_R} + \log \sigma_{z_I} - 2 \log |h_i^H w_i|$$

$$= -V_1(E\{G_1(\bar{z}_R)\}) - V_2(E\{G_2(\bar{z}_I)\}) + C_i,$$

where $h_i$ is a unit Euclidian length vector that is perpendicular to all the row vectors of $W$ except $w_i$, and $C_i$ is a constant term with respect to $w_i$.

The gradient update rule is given by

$$\frac{\partial J_i(w_i)}{\partial w_i} = \frac{1}{2} \frac{\partial \sigma_{z_R}^2}{\partial w_i} + \frac{1}{2} \frac{\partial \sigma_{z_I}^2}{\partial w_i} - \frac{h_i}{w_i^H h_i}$$

$$-v_{11}(E\{G_{11}(\bar{z}_{1R})\}) E\left\{g_{11}(\bar{z}_{1R}) \frac{\partial \bar{z}_{1R}}{\partial w_i^*}\right\}$$

$$-v_{12}(E\{G_{12}(\bar{z}_{1I})\}) E\left\{g_{12}(\bar{z}_{1I}) \frac{\partial \bar{z}_{1I}}{\partial w_i^*}\right\}.$$

The details of the derivations are given in Appendix B.

Since the update of the whitening filter contributes most to the CPU time, we also provide a computationally light version, complex entropy rate bound minimization-light (CERBM-L), by using the closed form approximation of $b$. By assuming $y$ is Gaussian distributed, the optimal $b$ is given by the eigenvector of $R_{y,K+1}^{-1} D_b$, where

$$R_{y,K+1}^{-1} = E\left\{y_{K+1} y_{K+1}^H\right\},$$

associated with the maximum eigenvalue.

5. EXPERIMENTAL RESULTS

In this section, we study the performances of proposed algorithms, CERBM and CERBM-L, in terms of the normalized interference to source ratio (ISR), which is given by $(1/N(N-1)) \sum_{i,j=1,i\neq j}^N E\{g_{ij}^2\}$, where $g_{ij}$ is the $ij$th entry of the global demixing matrix $G = WA$. All results are the average of 100 trials.

5.1. Performance of the new entropy rate estimator exploiting all three types of diversity

In order to show that the new method exploits all three types of diversity, we show the performance of the new entropy rate estimator using data that is generated by $s(t) = a s(t - 1) + z(t)$, where $a$ controls the sample dependence, and $z$ is an i.i.d. complex generalized Gaussian distributed (CGGD) process [28] with shape parameter $\beta$. The CGGD comprises a number of symmetric and unimodal distributions, from super-Gaussian $(0 < \beta < 1)$, Gaussian $(\beta = 1)$, to sub-Gaussian $(\beta > 1)$. The non-Gaussianity and non-circularity of $z$ are controlled by the shape parameter $\beta$ and $\rho = E\{z^2\}$, respectively. The entropy rate of $s$ equals to the entropy of $z$, since the coefficients of $s(t)$ and $z(t)$ are equal in the generative model, and $z$ is an i.i.d. process. The entropy of a CGGD random variable is given in [29], since a CGGD random variable can be considered as a bivariate GGD vector.

As observed in Fig.1, entropy rate $H_r$ decreases with increasing non-Gaussianity, i.e., when the CGGD shape parameter $\beta$ moves away from one, increasing diversity in terms of sample dependence, i.e., as the AR coefficient $a$ increases, and/or increasing diversity in terms of non-circularity, i.e., as the coefficient $\rho$ increases. Furthermore, we observe that the new entropy rate estimator, $\hat{H}_r$, accounts for all three types of diversity effectively in terms of approaching the true entropy rate $H_r$. The entropy estimator, $\hat{H}$, provides better estimation when there is no sample dependence, since it assumes that data is i.i.d. CGGD, which matches to the data exactly. However, it ignores sample dependence and thus its performance is degraded relative to $\hat{H}_r$ when samples are dependent.

5.2. CERBM performance for communications signal

In order to show the effectiveness of CERBM for a richer class of sources, their performances are compared with some widely used complex BSS algorithms in the separation of artificial mixtures of quadrature amplitude modulation (QAM) data. In this experiment, we generate sources by fourth-order moving-average (MA) models, with random coefficients, driven by i.i.d. QAM sources with order $2^n$ for the $n$th source. We vary the number of samples, and fix the number of sources and whitening filter length to be 5 and 4, respectively. From Fig.2, we observe that CERBM perform the best among those algorithms in terms of percentage of failures and normalized ISR. But CERBM is the most time consuming one due to
the updating of whitening filter coefficients. CERBM-L is faster than CERBM, since it uses a closed form approximation for the whitening filter coefficients.

![Figure 1](image1.png)  
**Fig. 1:** Performances of entropy rate estimator in terms of three types of diversity. Source is generated by first order AR model driven by an i.i.d. CGGD process. The true entropy rate, entropy rate estimated by the new estimator, and entropy estimated by assuming samples are i.i.d. CGGD are given by \( H_r, \hat{H}_r, \) and \( \hat{H} \), respectively. Note how the true entropy rate, \( H_r \), changes by varying the three types of diversity, and the entropy rate estimator, \( \hat{H}_r \), exploits all three.

![Figure 2](image2.png)  
**Fig. 2:** Three performance measures as function of samples size. Five sources are generated by fourth order MA model driven by i.i.d. QAM processes.

6. CONCLUSION

We propose an effective entropy rate estimator for complex-valued random process by using a flexible density matching method and assuming that sources can be whitened by widely linear filters. Based on the new entropy rate estimator, we introduce a new CICA algorithm, CERBM, that makes use of all three types of diversity: non-Gaussianity, sample dependency, and noncircularity. Simulation results show that the new entropy rate estimator accounts for three types of diversity and the effectiveness of CERBM. Other than these three types of diversity, it would be also interesting to consider nonstationarity as an additional property.

A. GRADIENT UPDATE RULE FOR WHITENING FILTER

Let \( \hat{R}_{z(K+1)} = E \left\{ Y_{(K+1)}Y_{(K+1)}^H \right\} \) and \( \hat{R}_{z(K+1)} = E \left\{ Y_{(K+1)}{Y}_{(K+1)}^\top \right\} \).

\[
\frac{\partial \log \sigma_{z_K}}{\partial \mathbf{b}^*} = \frac{1}{4 \sigma_{z_R}^2} \left( \hat{R}_{(K+1)} \mathbf{b}^* + \hat{R}_{(K+1)} \mathbf{b} \right)
\]

\[
\frac{\partial \log \sigma_{z_1}}{\partial \mathbf{b}^*} = -\frac{1}{4 \sigma_{z_1}^2} \left( \hat{R}_{(K+1)} \mathbf{b}^* - \hat{R}_{(K+1)} \mathbf{b} \right)
\]

\[
\frac{\partial V_1 \{ E \{ G_1(\tilde{z}_R) \} \} }{\partial \mathbf{b}^*} = v_1 \{ E \{ G_1(\tilde{z}_R) \} \} \left( \frac{E \{ g_1(\tilde{z}_R)Y_{(K+1)} \} }{2 \sigma_{z_R}} \right)
\]

\[
- \frac{E \{ g_1(\tilde{z}_R)^2 \} }{4 \sigma_{z_R}^2} \hat{R}_{(K+1)} \mathbf{b}^* - \frac{E \{ g_1(\tilde{z}_R)\tilde{z}_R \} }{4 \sigma_{z_R}^2} \hat{R}_{(K+1)} \mathbf{b}
\]

\[
\frac{\partial V_2 \{ E \{ G_2(\tilde{z}_1) \} \} }{\partial \mathbf{b}^*} = v_2 \{ E \{ G_2(\tilde{z}_1) \} \} \left( \frac{E \{ g_2(\tilde{z}_1)Y_{(K+1)} \} }{2 \sigma_{z_1}} \right)
\]

\[
+ \frac{E \{ g_2(\tilde{z}_1)^2 \} }{4 \sigma_{z_1}^2} \hat{R}_{(K+1)} \mathbf{b}^* - \frac{E \{ g_2(\tilde{z}_1)\tilde{z}_1 \} }{4 \sigma_{z_1}^2} \hat{R}_{(K+1)} \mathbf{b}
\]

where \( v_1 \) and \( v_2 \) are the derivatives of \( V_1 \) and \( V_2 \), respectively, and \( g_1 \) and \( g_2 \) are the derivatives of \( G_1 \) and \( G_2 \), respectively.

B. GRADIENT UPDATE RULE FOR CERBM

\[
\frac{\partial z_{1R}}{\partial \mathbf{w}_i} = \frac{1}{2} \left( X_{(K+1)}^* \mathbf{p}_i + X_{(K+1)} \mathbf{q}_i \right)
\]

\[
\frac{\partial z_{1I}}{\partial \mathbf{w}_i} = \frac{1}{2 j} \left( X_{(K+1)}^* \mathbf{p}_i - X_{(K+1)} \mathbf{q}_i \right)
\]

\[
\frac{\partial \sigma_{z_R}^2}{\partial \mathbf{w}^*} = \frac{1}{2} \left( E \left\{ zX_{(K+1)} \right\} + E \left\{ z^*X_{(K+1)} \right\} \right) \left( \mathbf{p}^* + \mathbf{q} \right)
\]

\[
\frac{\partial \sigma_{z_I}^2}{\partial \mathbf{w}^*} = -\frac{1}{2} \left( E \left\{ zX_{(K+1)} \right\} - E \left\{ z^*X_{(K+1)} \right\} \right) \left( \mathbf{q}^* - \mathbf{q} \right)
\]
C. REFERENCES


