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

Geng-Shen Fu, Ronald Phlypo, Matthew Anderson, Xi-Lin Li, and Tülay 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 properities 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. Noncircular 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 $\underline{\mathbf{x}} = [\mathbf{x}^\top, \mathbf{x}^H]^\top$ be the complex augmented vector, where superscript \top and H denote the transpose and Hermitian, respectively. The second-order statistics are given by the following augmented covariance matrix

$$\underline{\mathbf{R}} \triangleq E\left\{\underline{\mathbf{x}} \ \underline{\mathbf{x}}^{\mathrm{H}}\right\} = \begin{bmatrix} \mathbf{R} & \mathbf{R} \\ \widetilde{\mathbf{R}}^{*} & \mathbf{R}^{*} \end{bmatrix}$$

where E denotes expectation, superscript * denotes complex conjugate, and $\mathbf{R} \triangleq E \{\mathbf{x}\mathbf{x}^{\mathrm{H}}\}\$ and $\widetilde{\mathbf{R}} \triangleq E \{\mathbf{x}\mathbf{x}^{\top}\}\$

This work was supported by the NSF grant NSF-CIF 1117056.

are called the covariance and pseudo-covariance (or complementary covariance) matrix, respectively.

The probability density function (PDF) and entropy rate of \mathbf{x} are defined as $p(\mathbf{x}) \triangleq p(\mathbf{x}_R, \mathbf{x}_I)$ and $H_r(\mathbf{x}) \triangleq$ $H_r(\mathbf{x}_R, \mathbf{x}_I) = \lim_{T\to\infty} E\{-\log p(\mathbf{x}_R, \mathbf{x}_I)\}/T$, respectively. The random vector \mathbf{x} is called second-order circular or proper if pseudo-covariance matrix $\widetilde{\mathbf{R}} = \mathbf{0}$ and strictly circular, or circular, if \mathbf{x} and $\mathbf{x}e^{j\theta}$ 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 $\mathbf{s}(t) = [s_1(t), \ldots, s_N(t)]^\top \in \mathbb{C}^N$ be mixed through an $N \times N$ nonsingular mixing matrix $\mathbf{A} \in \mathbb{C}^{N \times N}$ so that we obtain the mixtures $\mathbf{x}(t) = [x_1(t), \ldots, x_N(t)]^\top$.

$$\mathbf{x}(t) = \mathbf{As}(t), 1 \le t \le T,$$

where t is the discrete sample index. The mixtures are separated as $\mathbf{y}(t) = \mathbf{W}\mathbf{x}(t)$, where $\mathbf{W} \triangleq [\mathbf{w}_1, \dots, \mathbf{w}_N]^{\mathrm{H}}$ is the demixing matrix, which is the quantity to be estimated for ICA, and $\mathbf{y}(t) \triangleq [y_1(t), \dots, y_N(t)]^{\mathrm{T}}$. This can also be written in matrix form, $\mathbf{Y} = \mathbf{W}\mathbf{X} = \mathbf{W}\mathbf{A}\mathbf{S}$, where $\mathbf{S} \triangleq [\mathbf{s}_1, \dots, \mathbf{s}_N]^{\mathrm{T}} \in \mathbb{C}^{N \times T}$, and $\mathbf{s}_i \in \mathbb{C}^T$. The same definition holds for \mathbf{x} and \mathbf{y} .

A natural cost for achieving the separation of these independent sources is the mutual information rate $\mathcal{I}_r(y_1; \ldots; y_N) = \sum_{i=1}^N H_r(y_i) - \log \det(\mathbf{W}\mathbf{W}^{\mathrm{H}}) - H_r(\mathbf{x})$ 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 $\mathbf{x}, H_r(\mathbf{x})$, is a constant with respect to \mathbf{W} . Hence, the cost function is given by:

$$\mathcal{J}_{r}(y_{1};\ldots;y_{N}) = \sum_{i=1}^{N} H_{r}(y_{i}) - 2\log|\det(\mathbf{W})|, \quad (1)$$

since $\det(\mathbf{WW}^{\mathrm{H}}) = |\det(\mathbf{W})|^2$. 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, i.e., there exists a b such that

$$\begin{split} \boldsymbol{z}(t) &= \mathbf{b}^{\mathrm{H}} \underline{\mathbf{y}}_{(K+1)}(t) = \mathbf{p}^{\mathrm{H}} \mathbf{y}_{(K+1)}(t) + \mathbf{q}^{\mathrm{H}} \mathbf{y}_{(K+1)}^{*}(t), \\ \text{where } \mathbf{p} &= [p_{0}, \dots, p_{K}]^{\top} \in \mathbb{C}^{K+1}, \mathbf{q} = [q_{0}, \dots, q_{K}]^{\top} \in \mathbb{C}^{K+1}, \\ \mathbf{b} &= [\mathbf{p}^{\top}, \mathbf{q}^{\top}]^{\top}, \mathbf{y}_{(K+1)}(t) = [y(t), \dots, y(t-K)]^{\top}, \\ \mathbf{b} &= [\mathbf{p}^{\top}, \mathbf{q}^{\top}]^{\top}, \mathbf{y}_{(K+1)}(t) = [y(t), \dots, y(t-K)]^{\top}. \end{split}$$

output process z will then be an i.i.d. process. We can always scale the whitening filter 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 b can be obtained by solving the following optimization problem:

$$\min_{\mathbf{b}} H(z), \text{ s.t. } \left| |p_0|^2 - |q_0|^2 \right| = 1,$$
(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

$$H(z) \le \log(\sigma_{z_R} \sigma_{z_I}) + H(\bar{z}_R) + H(\bar{z}_I), \qquad (3)$$

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

$$L_{\mathbf{b}}(\mathbf{b},\lambda) = \log(2\pi e \sigma_{z_R} \sigma_{z_I}) - V_1 \left(E \left\{ G_1(\bar{z}_R) \right\} \right) -V_2 \left(E \left\{ G_2(\bar{z}_I) \right\} \right) + \lambda \left(\mathbf{b}^H \mathbf{D}_b \mathbf{b} - 1 \right), (4)$$

where

$$\mathbf{D}_b = \operatorname{diag}\left(1, \underbrace{0, \dots, 0}_{K}, -1, \underbrace{0, \dots, 0}_{K}\right)$$

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 (proof omitted due to space constraints). 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-Guassianity, sample dependency, and non-circularity. Using the new entropy rate estimator (3), the entropy rate $H_r(y_i)$ 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 $\mathcal{J}_r(\mathbf{W})$ with respect to the demixing matrix \mathbf{W} , we use a decoupling procedure [26, 27] to divide the problem into minimizing $\mathcal{J}_r(\mathbf{W})$ with respect to each of the row vectors $\mathbf{w}_i, i = 1, \dots, N$. We can then write the cost as a function of only \mathbf{w}_i , which is

$$\begin{aligned} \mathcal{J}_{i}(\mathbf{w}_{i}) &= H(z_{i}) - 2\log|\mathbf{h}_{i}^{\mathrm{H}}\mathbf{w}_{i}| + C_{i} \\ &\leq \log(2\pi e) + \log\sigma_{z_{R}} + \log\sigma_{z_{I}} - 2\log|\mathbf{h}_{i}^{\mathrm{H}}\mathbf{w}_{i}| \\ &- V_{1}\left(E\left\{G_{1}(\bar{z}_{R})\right\}\right) - V_{2}\left(E\left\{G_{2}(\bar{z}_{I})\right\}\right) + C_{i}, \end{aligned}$$

where \mathbf{h}_i is a unit Euclidian length vector that is perpendicular to all the row vectors of \mathbf{W} except \mathbf{w}_i , and C_i is a constant term with respect to \mathbf{w}_i .

The gradient update rule is given by

$$\begin{split} \frac{\partial J_i(\mathbf{w}_i)}{\partial \mathbf{w}_i^*} &= \frac{1}{2\sigma_{z_{iR}}^2} \frac{\partial \sigma_{z_{iR}}^2}{\partial \mathbf{w}^*} + \frac{1}{2\sigma_{z_{iI}}^2} \frac{\partial \sigma_{z_{iI}}^2}{\partial \mathbf{w}^*} - \frac{\mathbf{h}_i}{\mathbf{w}_i^{\mathrm{H}} \mathbf{h}_i} \\ &- v_{i1} \left(E \left\{ G_{i1}(\bar{z}_{iR}) \right\} \right) E \left\{ g_{i1}(\bar{z}_{iR}) \frac{\partial \bar{z}_{iR}}{\partial \mathbf{w}_i^*} \right\} \\ &- v_{i2} \left(E \left\{ G_{i2}(\bar{z}_{iI}) \right\} \right) E \left\{ g_{i2}(\bar{z}_{iI}) \frac{\partial \bar{z}_{iI}}{\partial \mathbf{w}_i^*} \right\}. \end{split}$$

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 minimizationlight (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 $\underline{\mathbf{R}}_{\mathbf{y}_{(K+1)}}^{-1} \mathbf{D}_{b}$, where $\underline{\mathbf{R}}_{\mathbf{y}_{(K+1)}} = E\left\{\underline{\mathbf{y}}_{(K+1)}\underline{\mathbf{y}}_{(K+1)}^{\mathrm{H}}\right\}$, 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 $\mathbf{G} = \mathbf{W}\mathbf{A}$. 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) = as(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 β . 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 noncircularity of z are controled by the shape parameter β 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 GGD shape parameter β 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 noncircularity, i.e., as the coefficient ρ 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



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 , H_r , and H, respectively. Note how the true entropy rate, H_r , changes by varying the three types of diversity, and the entropy rate estimator, H_r , exploits all three.

the updating of whitening filter coefficients. CERBM-L is faster than CERBM, since it uses a closed form approximation for the whitening filter coefficients.



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
$$\underline{\mathbf{R}}_{(K+1)} = E\left\{\underline{\mathbf{y}}_{(K+1)}\underline{\mathbf{y}}_{(K+1)}^{\mathrm{H}}\right\}$$
 and $\underline{\widetilde{\mathbf{R}}}_{(K+1)} = E\left\{\underline{\mathbf{y}}_{(K+1)}\underline{\mathbf{y}}_{(K+1)}^{\mathrm{T}}\right\}$.
 $\frac{\partial \log \sigma_{z_{R}}}{\partial \mathbf{b}^{*}} = \frac{1}{4\sigma_{z_{R}}^{2}} \left(\underline{\widetilde{\mathbf{R}}}_{(K+1)}\mathbf{b}^{*} + \underline{\mathbf{R}}_{(K+1)}\mathbf{b}\right)$
 $\frac{\partial \log \sigma_{z_{I}}}{\partial \mathbf{b}^{*}} = \frac{-1}{4\sigma_{z_{I}}^{2}} \left(\underline{\widetilde{\mathbf{R}}}_{(K+1)}\mathbf{b}^{*} - \underline{\mathbf{R}}_{(K+1)}\mathbf{b}\right)$
 $\frac{\partial V_{1}\left\{E\left\{G_{1}(\bar{z}_{R})\right\}\right\}}{\partial \mathbf{b}^{*}} = v_{1}\left\{E\left\{G_{1}(\bar{z}_{R})\right\}\right\} \left(\frac{E\left\{g_{1}(\bar{z}_{R})\underline{\mathbf{y}}_{(K+1)}\right\}}{2\sigma_{z_{R}}}\right)$
 $-\frac{E\left\{g_{1}(\bar{z}_{R})\bar{z}_{R}\right\}}{4\sigma_{z_{R}}^{2}}\underline{\widetilde{\mathbf{R}}}_{(K+1)}\mathbf{b}^{*} - \frac{E\left\{g_{1}(\bar{z}_{R})\bar{z}_{R}\right\}}{4\sigma_{z_{R}}^{2}}\underline{\mathbf{R}}_{(K+1)}\mathbf{b}\right)$
 $\frac{\partial V_{2}\left\{E\left\{G_{2}(\bar{z}_{I})\right\}\right\}}{\partial \mathbf{b}^{*}} = v_{2}\left\{E\left\{G_{2}(\bar{z}_{I})\right\}\right\} \left(\frac{E\left\{g_{2}(\bar{z}_{I})\underline{\mathbf{y}}_{(K+1)}\right\}}{2j\sigma_{z_{I}}}\right)$
 $+\frac{E\left\{g_{2}(\bar{z}_{I})\bar{z}_{I}\right\}}{4\sigma_{z_{I}}^{2}}\underline{\widetilde{\mathbf{R}}}_{(K+1)}\mathbf{b}^{*} - \frac{E\left\{g_{2}(\bar{z}_{I})\bar{z}_{I}\right\}}{4\sigma_{z_{I}}^{2}}\underline{\mathbf{R}}_{(K+1)}\mathbf{b}\right),$

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

$$\begin{split} \frac{\partial z_{iR}}{\partial \mathbf{w}_i^*} &= \frac{1}{2} \left(\mathbf{X}_{(K+1)} \mathbf{p}_i^* + \mathbf{X}_{(K+1)} \mathbf{q}_i \right) \\ &\qquad \frac{\partial z_{iI}}{\partial \mathbf{w}_i^*} = \frac{1}{2j} \left(\mathbf{X}_{(K+1)} \mathbf{p}_i^* - \mathbf{X}_{(K+1)} \mathbf{q}_i \right) \\ \frac{\partial \sigma_R^2}{\partial \mathbf{w}^*} &= \frac{1}{2} \left(E \left\{ z \mathbf{X}_{(K+1)} \right\} + E \left\{ z^* \mathbf{X}_{(K+1)} \right\} \right) (\mathbf{p}^* + \mathbf{q}) \\ \frac{\partial \sigma_I^2}{\partial \mathbf{w}^*} &= \frac{-1}{2} \left(E \left\{ z \mathbf{X}_{(K+1)} \right\} - E \left\{ z^* \mathbf{X}_{(K+1)} \right\} \right) (\mathbf{q}^* - \mathbf{q}) \end{split}$$

 ∂

C. REFERENCES

- T. Adalı, P. Schreier, and L. Scharf, "Complex-valued signal processing: The proper way to deal with impropriety," *Signal Processing, IEEE Transactions on*, vol. 59, no. 11, pp. 5101–5125, 2011.
- [2] P. Comon and C. Jutten, Handbook of Blind Source Separation: Independent Component Analysis and Applications, 1st ed. Academic Press, 2010.
- [3] T. Adalı and V. D. Calhoun, "Complex ICA of brain imaging data," *Signal Processing Magazine*, *IEEE*, vol. 24, no. 5, pp. 136–139, 2007.
- [4] T. Adalı, M. Anderson, and G.-S. Fu, "IVA and ICA: Use of diversity in independent decompositions," in *Signal Processing Conference (EUSIPCO), 2012 Proceedings of the 20th European*, Aug. 2012, pp. 61–65.
- [5] E. Moreau and T. Adalı, Blind Identification and Separation of Complex-valued Signals. London, UK and Hoboken, NJ: Wiley-ISTE, 2013.
- [6] B. Loesch and B. Yang, "Cramér-Rao bound for circular and noncircular complex independent component analysis," *Signal Processing, IEEE Transactions on*, vol. 61, no. 2, pp. 365–379, Jan. 2013.
- [7] L. De Lathauwer and B. De Moor, "On the blind separation of non-circular sources," in *EUSIPCO-02*, Toulouse, France, sept. 2002.
- [8] J. Eriksson and V. Koivunen, "Complex-valued ICA using second order statistics," in *Machine Learning for Signal Processing*, 2004. Proceedings of the 2004 14th IEEE Signal Processing Society Workshop, 2004, pp. 183–192.
- [9] E. Ollila and V. Koivunen, "Complex ICA using generalized uncorrelating transform," *Signal Process.*, vol. 89, no. 4, pp. 365–377, Apr. 2009.
- [10] J.-F. Cardoso and A. Souloumiac, "Blind beamforming for non-Gaussian signals," *Radar and Signal Processing*, *IEE Proceedings F*, vol. 140, no. 6, pp. 362–370, Dec. 1993.
- [11] E. Bingham and A. Hyvärinen, "A fast fixed-point algorithm for independent component analysis of complex valued signals." *International journal of neural systems*, vol. 10, no. 1, pp. 1–8, Feb. 2000.
- [12] P. Smaragdis, "Blind separation of convolved mixtures in the frequency domain," *Neurocomputing*, vol. 22, no. 1-3, pp. 21–34, Nov. 1998.
- [13] J. Anemüller, T. J. Sejnowski, and S. Makeig, "Complex independent component analysis of frequency-domain electroencephalographic data," *Neural Netw.*, vol. 16, no. 9, pp. 1311–1323, Nov. 2003.
- [14] M. Novey and T. Adalı, "On extending the complex FastICA algorithm to noncircular sources," *Signal Processing, IEEE Transactions on*, vol. 56, no. 5, pp. 2148–2154, 2008.
- [15] S. C. Douglas, "Fixed-point algorithms for the blind separation of arbitrary complex-valued non-Gaussian signal mixtures," *EURASIP J. Appl. Signal Process.*, vol. 2007, no. 1, pp. 83–83, Jan. 2007.

- [16] H. Li and T. Adalı, "A class of complex ICA algorithms based on the kurtosis cost function," *Neural Networks*, *IEEE Transactions on*, vol. 19, no. 3, pp. 408–420, 2008.
- [17] X.-L. Li and T. Adali, "Complex independent component analysis by entropy bound minimization," *Circuits and Systems I: Regular Papers, IEEE Transactions on*, vol. 57, no. 7, pp. 1417–1430, Jul. 2010.
- [18] T. Adalı, H. Li, M. Novey, and J.-F. Cardoso, "Complex ICA using nonlinear functions," *Signal Processing, IEEE Transactions on*, vol. 56, no. 9, pp. 4536–4544, 2008.
- [19] H. Sawada, R. Mukai, S. Araki, and S. Makino, "A polarcoordinate based activation function for frequency domain blind source separation," in *Proceedings of the 3rd international conference on independent component analysis and blind signal separation*, 2001, pp. 663–668.
- [20] J.-F. Cardoso and T. Adalı, "The maximum likelihood approach to complex ICA," in Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings. 2006 IEEE International Conference on, vol. 5, 2006, pp. V–V.
- [21] M. Novey and T. Adalı, "Complex ICA by negentropy maximization," *Neural Networks, IEEE Transactions on*, vol. 19, no. 4, pp. 596–609, 2008.
- [22] X.-L. Li and T. Adalı, "Blind separation of noncircular correlated sources using Gaussian entropy rate," *Signal Processing, IEEE Transactions on*, vol. 59, no. 6, pp. 2969–2975, Jun. 2011.
- [23] G.-S. Fu, R. Phlypo, M. Anderson, X.-L. Li, and T. Adalı, "Algorithms for Markovian source separation by entropy rate minimization," in *Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference on*, 2013, pp. 3248–3252.
- [24] X.-L. Li and T. Adalı, "Independent component analysis by entropy bound minimization," *Signal Processing*, *IEEE Transactions on*, vol. 58, no. 10, pp. 5151–5164, Oct. 2010.
- [25] T. Adalı and S. Haykin, Adaptive Signal Processing: Next Generation Solutions, ser. Adaptive and Learning Systems for Signal Processing, Communications and Control Series. Wiley, 2010.
- [26] X.-L. Li and X.-D. Zhang, "Nonorthogonal joint diagonalization free of degenerate solution," *Signal Processing*, *IEEE Transactions on*, vol. 55, no. 5, pp. 1803–1814, May 2007.
- [27] M. Anderson, X.-L. Li, P. Rodriguez, and T. Adalı, "An effective decoupling method for matrix optimization and its application to the ICA problem," in Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on, Mar. 2012, pp. 1885–1888.
- [28] M. Novey, T. Adalı, and A. Roy, "A complex generalized Gaussian distribution – characterization, generation, and estimation," *Signal Processing, IEEE Transactions on*, vol. 58, no. 3, pp. 1427–1433, 2010.
- [29] G. Aulogiaris and K. Zografos, "A maximum entropy characterization of symmetric Kotz type and Burr multivariate distributions," *TEST*, vol. 13, pp. 65–83, 2004.