A METHOD BASED ON L-BFGS TO SOLVE CONSTRAINED COMPLEX-VALUED ICA

Anh H. T. Nguyen[1]*, V. G. Reju*[2]*, and Andy W. H. Khong*[1]

[1]Nanyang Technological University, Singapore
[2]Cochin University of Science and Technology, Kochi, Kerala, India
Email: nguyenha001@e.ntu.edu.sg, reju@cusat.ac.in, andykhong@ntu.edu.sg.

ABSTRACT

Complex-valued independent component analysis (ICA) is a celebrated method in blind separation of complex-valued signals. In this paper, we propose to transform the constrained optimization problems of complex-valued ICA into unconstrained optimization problems which can be solved by limited-memory Broyden–Fletcher–Goldfarb–Shanno update (L-BFGS). As opposed to previous approaches, the proposed method does not apply any restriction on the Hessian matrix of ICA cost function. It can separate mixed sub-Gaussian, super-Gaussian, circular, and non-circular sources. Simulations show promising results.

Index Terms: complex-valued ICA, L-BFGS, blind source separation

1. INTRODUCTION

Complex-valued ICA is an unsupervised algorithm which has been applied to a wide range of applications across different fields, particularly biomedical signal processing [1], [2] and blind separation of convolutive audio mixtures [3]–[5]. A notable complex-valued ICA approach is the use of nonlinear functions to achieve one of several related objectives such as maximization of non-Gaussianity, minimization of negentropy, maximum likelihood, and minimization of mutual information [6]–[11].

As predicted by the Cramer-Rao lower bound (CRLB), ICA cost function becomes increasingly flat at the underlying mixing matrix when the source distributions approach Gaussian distribution. This can be seen from the relationship between the Fisher Information matrix and the Hessian matrix of the negative log likelihood. Therefore, it is favorable to use second-order optimization methods which exploit the curvature of the cost functions to escape plateau regions. Indeed, fixed-point iteration, a quasi-Newton method, is the most common implementation of ICA [12]. However, the Hessian approximation in fixed-point ICA techniques is restrictive and may not reflect the true curvature of ICA objectives well. For a better approximation of the Hessian matrix, L-BFGS-based ICA have been proposed [13], [14]. These methods are not suitable for constrained ICA. In [15], the authors propose a real-valued orthogonally constrained ICA by performing L-BFGS updates on the Stiefel manifold. Nevertheless, the algorithm in [15] requires the estimate of the sign of the kurtosis – a highly noisy statistics especially when the sources are close to Gaussian. It must also reset L-BFGS whenever the kurtosis of any demixed signal changes its sign.

In this work, we propose to formulate constrained complex-valued ICA as an unconstrained optimization problem with respect to auxiliary variables and solve this problem with vanilla L-BFGS [16]. The proposed method does not apply any restriction on the joint Hessian. This will reflect the true Hessian more accurately than fixed-point iteration. Coupled with entropy bound minimization criterion [10], the proposed algorithm can separate mixtures of sub-Gaussian, super-Gaussian, circular, and non-circular sources.

2. BACKGROUND

2.1. Complex vector calculus

We use $E\{\cdot\}$, $(\cdot)^T$, $(\cdot)^*$, and $(\cdot)^H$ to denote expectation, transposition, conjugation, and conjugate transposition respectively. The identity matrix in $\mathbb{C}^{N\times N}$ is denoted by $\mathbf{I}_N$. The imaginary unit is denoted by $\jmath = \sqrt{-1}$. The real part and imaginary part of a complex vector $\mathbf{z}$ are respectively denoted by $\Re\{\mathbf{z}\}$ and $\Im\{\mathbf{z}\}$. The L_2-norm of $\mathbf{z}$ is denoted by $\|\mathbf{z}\|$. For an arbitrarily chosen complex-valued random variable $z = \Re\{z\} + \jmath\Im\{z\}$, the probability distribution and entropy of z are defined by $f_Z(z) \triangleq f_Z(\Re\{z\}, \Im\{z\})$ and $H(z) \triangleq H(\Re\{z\}, \Im\{z\}) = -E\{\log f_Z(\Re\{z\}, \Im\{z\})\}$.

The cost functions of complex-valued ICA are often real-valued functions of complex-valued arguments. Since these functions are non analytic, to optimize them, we can parameterize them in two equivalent forms – real-composite form and complex-augmented form [17]–[19], where the latter is usually known as Wirtinger calculus. For a complex vector $\mathbf{w} \in \mathbb{C}^N$, one can respectively define

This work was supported within the Delta-NTU Corporate Lab for Cyber-Physical Systems with funding support from Delta Electronics Inc and the National Research Foundation (NRF) Singapore under the Corp Lab@ University Scheme (Ref. SLE-RP5) at Nanyang Technological University, Singapore.

its real-composite vector and complex-augmented vector as follows: $\mathbf{w}_{\mathcal{R}} = [\Re\{\mathbf{w}\}^T, \Im\{\mathbf{w}\}^T]^T$, and $\mathbf{w}_{\mathcal{C}} = [\mathbf{w}^T, \mathbf{w}^H]^T$. The mapping from $\mathbf{w}_{\mathcal{R}}$ to $\mathbf{w}_{\mathcal{C}}$ is a one-to-one linear map $\mathbf{w}_{\mathcal{C}} = \mathbf{T}_N \mathbf{w}_{\mathcal{R}}$, where $\mathbf{T}_N = \begin{bmatrix} \mathbf{I}_N & \jmath\mathbf{I}_N \\ \mathbf{I}_N & -\jmath\mathbf{I}_N \end{bmatrix}$ and $\mathbf{T}_N^{-1} = \frac{1}{2}\mathbf{T}_N^H$. Due to linearity of differentiation, for a generic vector function $J(\mathbf{w}) : \mathbb{C}^N \mapsto \mathbb{C}^M$, we have

$$\frac{\partial J}{\partial \mathbf{w}_{\mathcal{C}}} = \tfrac{1}{2}\mathbf{T}_N^* \frac{\partial J}{\partial \mathbf{w}_{\mathcal{R}}} = \frac{1}{2}\begin{bmatrix} \frac{\partial J}{\partial \Re\{\mathbf{w}\}} - \jmath \frac{\partial J}{\partial \Im\{\mathbf{w}\}} \\ \frac{\partial J}{\partial \Re\{\mathbf{w}\}} + \jmath \frac{\partial J}{\partial \Im\{\mathbf{w}\}} \end{bmatrix}, \quad (1)$$

$$\frac{\partial J}{\partial \mathbf{w}_{\mathcal{R}}} = \mathbf{T}_N^T \frac{\partial J}{\partial \mathbf{w}_{\mathcal{C}}} = 2\begin{bmatrix} \Re\{\frac{\partial J}{\partial \mathbf{w}^*}\} \\ \Im\{\frac{\partial J}{\partial \mathbf{w}^*}\} \end{bmatrix}. \quad (2)$$

As $\mathbf{w} = \Re\{\mathbf{w}\} + \jmath\Im\{\mathbf{w}\}$, (1) implies that $\frac{\partial \mathbf{w}}{\partial \mathbf{w}^*} = \frac{\partial \mathbf{w}^*}{\partial \mathbf{w}} = 0$. This means that analyzing $J(\mathbf{w}_{\mathcal{C}})$ is more convenient than analyzing $J(\mathbf{w}_{\mathcal{R}})$ since one can treat $\mathbf{w}$ as a constant when finding derivative w.r.t. $\mathbf{w}^*$ and vice versa. In addition, (2) allows us to find $\frac{\partial J}{\partial \mathbf{w}_{\mathcal{R}}}$ conveniently instead of complicated direct computation. When the co-domain of $J(\mathbf{w})$ is $\mathbb{R}$, it can be shown that the steepest descend directions of $J(\mathbf{w}_{\mathcal{R}})$ and $J(\mathbf{w}_{\mathcal{C}})$ are, respectively, given by $-\frac{\partial J(\mathbf{w}_{\mathcal{R}})}{\partial \mathbf{w}_{\mathcal{R}}}$ and $-\frac{\partial J(\mathbf{w}_{\mathcal{C}})}{\partial \mathbf{w}_{\mathcal{C}}^*}$.

2.2. Complex-valued ICA

We suppose that the data follows the linear model

$$\mathbf{x} = \mathbf{A}\mathbf{s}, \quad (3)$$

where $\mathbf{x} = [x_1, \ldots, x_N]^T$ is the vector of observed data, $\mathbf{A} \in \mathbb{C}^{N\times N}$ is the unknown invertible mixing matrix, N is the number of sources, and $\mathbf{s} = [s_1, \ldots, s_N]^T$ is the vector of unknown independent sources.

Let $\mathbf{W} = [\mathbf{w}_1, \ldots, \mathbf{w}_N]$ be the demixing matrix and let $\mathbf{y} = [y_1, \ldots, y_N]^T = \mathbf{W}^H\mathbf{x}$ be the demixed signals. To estimate $\mathbf{s}$, one can minimize the mutual information of demixed signals given by [6], [10]

$$\mathcal{I}(\mathbf{W}) = \textstyle\sum_i H(y_i) - 2\log|\det(\mathbf{W})| - C, \quad (4)$$

where $H(y_i)$ is the entropy of y_i, and C is a term which does not depend on $\mathbf{W}$. Since $\mathcal{I}(\mathbf{W})$ is unbounded below since the second term can be arbitrarily small, certain restriction on $\mathbf{W}$ should be applied. When the data are whitened, one can constrain $\mathbf{W}$ to be unitary (i.e., $\mathbf{W}\mathbf{W}^H = \mathbf{W}^H\mathbf{W} = \mathbf{I}$). In this case, the second term of (4) can be ignored. A more relaxed restriction is to constrain each column of $\mathbf{W}$ to have unit norm so that $|\det(\mathbf{W})|$ is bounded above by 1.

In the so-called complex-valued ICA by entropy bound minimization (CEBM) [10], several upper bounds of $H(y_i)$ corresponding to some predetermined measuring functions are estimated based on maximum entropy principle. The tightest upper bound and corresponding nonlinear function(s) will subsequently be chosen to compute the value and the gradient of each $H(y_i)$. Using a decoupling trick [20], the demixing filters $\mathbf{w}_1, \ldots, \mathbf{w}_N$ are updated independently using projected gradient descending in CEBM. As projected gradient descending does not exploit the curvature information, this may lead to sub-optimal performance in flat regions of $\mathcal{I}(\mathbf{W})$. Although, fixed-point iteration can remedy this issue to certain extend, the update of $\mathbf{w}_1, \ldots, \mathbf{w}_N$ are also computed independently in this approach. Consequently, the Hessian of $\mathrm{vec}(\mathbf{W})$, a vector created by stacking all columns of $\mathbf{W}$, is restricted to be block diagonal in fixed-point iteration. This may also affect the estimation performance when the independent model does not hold exactly.

In the next section, we propose a second-order method based on L-BFGS without the above mentioned limitations.

3. COMPLEX-VALUED ICA BASED ON L-BFGS

For $\mathcal{I}(\mathbf{W})$ given in (4), suppose that the total entropy bound $\mathcal{H}(\mathbf{W}) = \sum_i H(\mathbf{w}_i^H\mathbf{x})$ and its gradient $\frac{\partial \mathcal{H}}{\partial \mathbf{W}^*}$ are estimated by numerical method in [10], the gradient direction of $\mathcal{I}(\mathbf{W})$ is then given by

$$\frac{\partial \mathcal{I}(\mathbf{W})}{\partial \mathbf{W}^*} = \frac{\partial \mathcal{H}(\mathbf{W})}{\partial \mathbf{W}^*} - \mathbf{W}^{-H}. \quad (5)$$

We consider two constrained optimization problems

$$\min_{\mathbf{W}} \ \mathcal{I}(\mathbf{W}), \text{ s.t. } \mathbf{W}\mathbf{W}^H = \mathbf{I}_M, \quad (6)$$

$$\min_{\mathbf{W}} \ \mathcal{I}(\mathbf{W}), \text{ s.t. } \|\mathbf{w}_1\| = \ldots = \|\mathbf{w}_N\| = 1. \quad (7)$$

For an arbitrary full rank matrix $\widetilde{\mathbf{W}} \in \mathbb{C}^{N\times N}$, let $\widehat{\mathbf{W}} = (\widetilde{\mathbf{W}}\widetilde{\mathbf{W}}^H)^{-0.5}\widetilde{\mathbf{W}}$ be the nearest unitary matrix of $\widetilde{\mathbf{W}}$ and $\overline{\mathbf{W}} = [\bar{\mathbf{w}}_1, \ldots, \bar{\mathbf{w}}_N]$ be the column normalized matrix of $\widetilde{\mathbf{W}}$, where $\bar{\mathbf{w}}_i = \tilde{\mathbf{w}}_i / \|\tilde{\mathbf{w}}_i\|$. It is important to note that $\widehat{\mathbf{W}}$ and $\overline{\mathbf{W}}$ are functions of $\widetilde{\mathbf{W}}$. By definition, $\widehat{\mathbf{W}}$ and $\overline{\mathbf{W}}$ always satisfy their respective constraints in (6) and (7). With these reparameterizations, we can indirectly solve (6) and (7) respectively by solving the following unconstrained optimization problems

$$\min_{\widetilde{\mathbf{W}}\in\mathbb{C}^{N\times N}} \ \mathcal{I}(\widehat{\mathbf{W}}) + \tfrac{\lambda_1}{2}\|\widetilde{\mathbf{W}}\|_F^2, \quad (8)$$

$$\min_{\widetilde{\mathbf{W}}\in\mathbb{C}^{N\times N}} \ \mathcal{I}(\overline{\mathbf{W}}) + \tfrac{\lambda_2}{2}\|\widetilde{\mathbf{W}}\|_F^2, \quad (9)$$

where λ_1 and λ_2 are two hyper-parameters and $\|\widetilde{\mathbf{W}}\|_F^2$ is the the Frobenius norm of $\widetilde{\mathbf{W}}$. It is noted that $\mathcal{I}(\widehat{\mathbf{W}})$ and $\mathcal{I}(\overline{\mathbf{W}})$ do not depend on $\|\widetilde{\mathbf{W}}\|_F^2$ as soon as $\widetilde{\mathbf{W}}$ is full rank. Therefore, given small values of λ_1 and λ_2, the second term of (8) and (9) will penalize $\widetilde{\mathbf{W}}$ which has large Frobenius norm. Nevertheless, for the exact minimization of $\mathcal{I}(\widehat{\mathbf{W}})$ and $\mathcal{I}(\overline{\mathbf{W}})$, one should use $\lambda_1 = \lambda_2 = 0$. Using Wirtinger matrix calculus, the gradient of the nested cost in (8) is given by [21]

$$\mathbf{K} = -\big(\boldsymbol{\Sigma}^{-1}\mathbf{U}^H\big(\tfrac{\partial \mathcal{I}(\widehat{\mathbf{W}})}{\partial \widehat{\mathbf{W}}^*}\big)\mathbf{V}\big) \oslash (\mathbf{1}_N\boldsymbol{\sigma}^T + \boldsymbol{\sigma}\mathbf{1}_N^T), \quad (10)$$

$$\frac{\partial \mathcal{I}(\widehat{\mathbf{W}})}{\partial \widetilde{\mathbf{W}}^*} = \mathbf{U}(\mathbf{K}^H + \mathbf{K})\boldsymbol{\Sigma}\mathbf{V}^H + \mathbf{U}\boldsymbol{\Sigma}^{-1}\mathbf{U}^H\big(\tfrac{\partial \mathcal{I}(\widehat{\mathbf{W}})}{\partial \widehat{\mathbf{W}}^*}\big), \quad (11)$$

Algorithm 1 Complex-valued ICA based on L-BFGS

Find initial matrix $\mathbf{W}_0$ using [9]
Solving (8) using L-BFGS, given initial matrix $\mathbf{W}_0$
$\widetilde{\mathbf{W}}_0 \leftarrow (\widetilde{\mathbf{W}}\widetilde{\mathbf{W}}^H)^{-0.5}\widetilde{\mathbf{W}}$
Solving (9) using L-BFGS, given initial matrix $\widetilde{\mathbf{W}}_0$
$\mathbf{W} \leftarrow [\frac{\widetilde{\mathbf{w}}_1}{\|\widetilde{\mathbf{w}}_1\|}, \ldots, \frac{\widetilde{\mathbf{w}}_N}{\|\widetilde{\mathbf{w}}_N\|}]$

where $\oslash$ is the element-wise matrix division, $\mathbf{U\Sigma V}^H$ denotes the singular value decomposition of $\widetilde{\mathbf{W}}$, $\boldsymbol{\sigma}$ is the diagonal of $\boldsymbol{\Sigma}$, and $\mathbf{1}_N \in \mathbb{C}^N$ is the all-ones column vector. Similarly, using Wirtinger calculus, the gradient w.r.t. each column of the nested cost given in (9) is

$$\begin{aligned}\frac{\partial \mathcal{I}(\overline{\mathbf{W}})}{\partial \widetilde{\mathbf{w}}_i^*} &= \frac{\partial \bar{\mathbf{w}}_i^*}{\partial \widetilde{\mathbf{w}}_i^*}\frac{\partial \mathcal{I}(\overline{\mathbf{W}})}{\partial \bar{\mathbf{w}}_i^*} + \frac{\partial \bar{\mathbf{w}}_i}{\partial \widetilde{\mathbf{w}}_i^*}\frac{\partial \mathcal{I}(\overline{\mathbf{W}})}{\partial \bar{\mathbf{w}}_i} \\ &= \left(\frac{\mathbf{I}}{\|\widetilde{\mathbf{w}}_i\|} - \frac{\widetilde{\mathbf{w}}_i\widetilde{\mathbf{w}}_i^H}{2\|\widetilde{\mathbf{w}}_i\|^3}\right)\frac{\partial \mathcal{I}(\overline{\mathbf{W}})}{\partial \bar{\mathbf{w}}^*} - \frac{\widetilde{\mathbf{w}}_i\widetilde{\mathbf{w}}_i^T}{2\|\widetilde{\mathbf{w}}_i\|^3}\frac{\partial \mathcal{I}(\overline{\mathbf{W}})}{\partial \bar{\mathbf{w}}} \\ &= \frac{1}{\|\widetilde{\mathbf{w}}_i\|}\left(\frac{\partial \mathcal{I}(\overline{\mathbf{W}})}{\partial \bar{\mathbf{w}}_i^*} - \bar{\mathbf{w}}_i \Re\left\{\bar{\mathbf{w}}_i^H \frac{\partial \mathcal{I}(\overline{\mathbf{W}})}{\partial \bar{\mathbf{w}}_i^*}\right\}\right). \end{aligned} \tag{12}$$

In the last step, since $\mathcal{I}(\overline{\mathbf{W}})$ is real-valued, we have used $(\frac{\partial \mathcal{I}}{\partial \bar{\mathbf{w}}})^* = \frac{\partial \mathcal{I}}{\partial \bar{\mathbf{w}}^*}$. The gradients, $\frac{\partial \mathcal{I}}{\partial \bar{\mathbf{w}}_i^*}$ and $\frac{\partial \mathcal{I}(\widehat{\mathbf{W}})}{\partial \widehat{\mathbf{W}}^*}$, are evaluated using (5).

Since the cost functions in (8) and (9) are unconstrained, one can vectorize these gradients and solve the respective cost functions in real-composite space or complex-augmented space. In this work, we choose to optimize in real-composite space. To do so, the complex-valued gradients are converted into real-composite space using (2). Here, the reparameterized objectives can be solved using any real-valued gradient-based solver. However, to address the performance reduction of ICA when the sources approach Gaussianity, we solve (8) and (9) using L-BFGS [16]. Note that an equivalent implementation in complex-augmented space is also possible using the complex-valued L-BFGS software accompanied by [17]. The proposed method is summarized in Algorithm 1. Our algorithm comprises three main steps. It is initialized using [9] with the contrast function of $G(y) = y^{1.25}$. After that, an ICA estimation is performed using (8), then the estimate is refined using (9). The refinement step are necessary because the demixing matrix is not exactly unitary due to imperfection of whitening [20]. The main advantage of our proposed method is that the approximate Hessian of $\mathrm{vec}(\widetilde{\mathbf{W}})$ based on L-BFGS is not limited to be block diagonal. The second advantage is that strong Wolfe line search of L-BFGS guarantees a sufficient decrease of the objective at each step, whereas there is no such guarantee in fixed-point iteration. Last but not least, the method can be applied for other complex-valued ICA objectives or any real ICA objectives with ease.

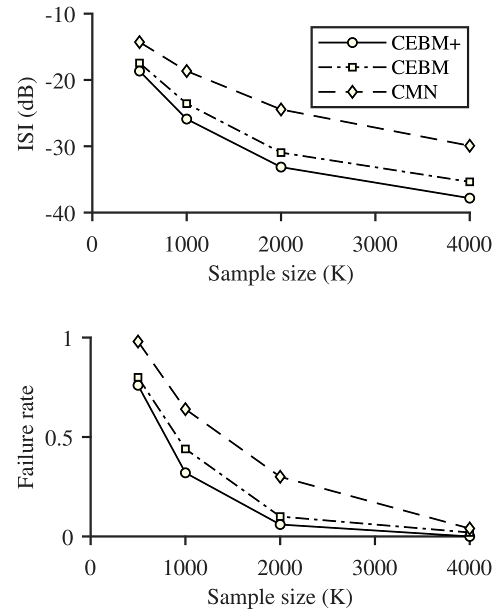

Fig. 1: Average ISI and failure rate as functions of the sample size.

4. SIMULATIONS

We compare the proposed complex-valued entropy bound minimization based on LBFGS (CEBM+), with CEBM [10] and complex-valued ICA by maximization of non-Gaussianity (CMN) [9]. Methods such as complex-valued FastICA [7] and complex-valued non-circular FastICA [8] are not included as they normally perform worse than CEBM and CMN. Complex-valued ICA by entropy rate bound minimization [11] is also not included because the focus of this paper is the optimization method. The proposed CEBM+ is implemented in MATLAB using minFunc toolbox [22] where the maximum number of iterations is 100 (40 iterations for (8) and 60 iterations for (9)). The algorithm is stopped if the directional gradient is smaller than 10^{-5}. We set hyperparameters λ_1 and λ_2 to 0.001.

We employ two performance indexes – the inter-symbol-interference (ISI) and the percentage of failed trials, both of which are commonly used in ICA literature. The ISI is defined as [6], [23]

$$\begin{aligned} I_A =& \frac{1}{2N(N-1)}\sum_{m=1}^N\left(\sum_{n=1}^N \frac{p_{mn}}{\max_l|p_{ml}|} - 1\right) \\ &+ \frac{1}{2N(N-1)}\sum_{m=1}^N\left(\sum_{n=1}^N \frac{p_{mn}}{\max_l|p_{ln}|} - 1\right), \end{aligned} \tag{13}$$

where $\mathbf{P} = [p_{ij}] = \mathbf{W}^H\mathbf{A}$. In a sense, I_A measures how close $\mathbf{W}^H\mathbf{A}$ is to the identity matrix. An algorithm is deemed to have failed if the matrix $\mathbf{P} = [p_{ij}]$ has two rows where the largest magnitude values are at the same column. Intuitively, the magnitude p_{ij} is the cosine-similarity of $\mathbf{w}_i$ and $\mathbf{a}_j$ (given that $\mathbf{W}$ and $\mathbf{A}$ are unitary due to prewhitening). If there are

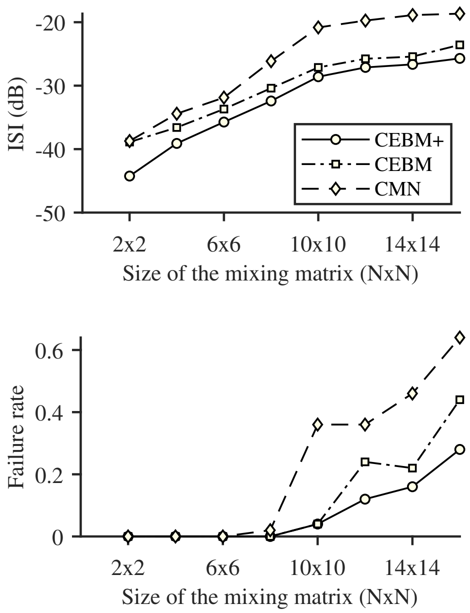

Fig. 2: Average ISI and failure rate as functions of the number of sources

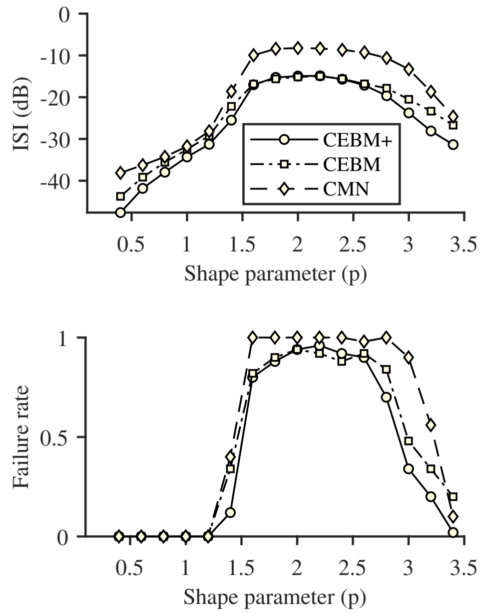

Fig. 3: Average ISI and failure rate as functions of the shape parameter ($p = 2$ is Gaussian).

two columns $\mathbf{w}_m$ and $\mathbf{w}_n$ that are both associated to the same column of $\mathbf{A}$, the algorithm should be considered as being failed. Lower ISI and lower failure percentage indicate better performance.

For each simulation, we repeat 100 trials, where for each trial, new $\mathbf{A}$ and $\mathbf{s}$ are generated. The mixed data are subsequently generated using (3). The real and imaginary part of each element of $\mathbf{A}$ are drawn from standard normal distribution. Each source s_i is drawn from zero-mean complex generalized Gaussian distribution [24] where two parameter p and ρ control the shape of the distribution and the circularity of the source, respectively. The circular coefficient ρ is defined by the correlation between $\Re\{s_i\}$ and $\Im\{s_i\}$ when variances of $\Re\{s_i\}$ and $\Im\{s_i\}$ are assumed to be 1. A source is complex Gaussian distributed if $p = 2$ and a source is circular if $\rho = 0$.

Figure 1 depicts the performance with respect to sample size. Here, we created sixteen sources ($N = 16$) with the shape parameters being evenly spaced between $0.5 \leq p \leq 3.5$. All sources are non-circular with ρ is selected in $(0, 1)$ randomly. This simulation shows that our proposed algorithm achieves lower sample complexity than CMN and CEBM. Figure 2 illustrates the performance as a function of the number of sources. Similar to previous simulation, we created sources with shape parameter that is uniformly spaced in the range $[0.5, 3.5]$. The sample size is fixed at 1000 samples. In this simulation, our proposed algorithm yields better estimates than CMN and CEBM. Note that, both our proposed method and CEBM share the same objective function. This highlights the benefits of joint optimization and the effectiveness of L-BFGS. Figure 3 shows the effect of shape parameter on the performance of the tested complex-valued ICA algorithms. In each trial, we generated sixteen i.i.d. complex generalized Gaussian sources. The sample size is 1000 and all the sources are non-circular with ρ is randomly selected in $(0, 1)$. The shape parameter p is varied from 0.4 to 3.4. As shown by this result, the estimation performance reduces whens the Gaussianity of the sources increase. In all tested cases, our algorithm generally performs better than the baseline algorithms for sources with low non-Gaussianity as seen from the failure rate. On average, the running time of CEBM+ is twice as much as CEBM due to different stopping conditions and the costly L-BFGS update.

Overall, the proposed method performs equally or higher than CMN and CEBM in most of the simulations. In particular, the proposed algorithm has lower sample complexity and can yields better estimate than CMN and CEBM when the source distributions are close to Gaussian.

5. CONCLUSIONS

We proposed an algorithm that employs L-BFGS to improve the estimation performance of complex-valued ICA. Since complex-valued ICA is naturally constrained, it is not feasible to use the L-BFGS algorithm directly. We address this issue by formulating complex-valued ICA as an unconstrained optimization problem. This allows us to approximate the joint Hessian of all demixing filter jointly using L-BFGS updating rule. As a result, the proposed algorithm outperforms that of its counterparts in many cases.

6. REFERENCES

[1] P. A. Rodriguez, M. Anderson, V. D. Calhoun, and T. Adali, "General nonunitary constrained ICA and its application to complex-valued fMRI data," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 3, pp. 922–929, 2015.

[2] T. Adali, M. Anderson, and G. Fu, "Diversity in independent component and vector analyses: Identifiability, algorithms, and applications in medical imaging," *IEEE Signal Process. Mag.*, vol. 31, no. 3, pp. 18–33, May 2014.

[3] J. Eriksson, A.-M. Seppola, and V. Koivunen, "Complex ICA for circular and non-circular sources," in *methods*, vol. 7. IEEE, 2005, pp. 1–4.

[4] N. Mitianoudis and M. Davies, "Audio source separation of convolutive mixtures," *IEEE Trans. Acoust. Speech Signal Process.*, vol. 11, no. 5, pp. 489–497, 2003.

[5] T. Nishikawa, H. Abe, H. Saruwatari, and K. Shikano, "Overdetermined blind separation for convolutive mixtures of speech based on multistage ICA using subarray processing," in *Proc. IEEE Int. Conf. Acoust. Speech, Signal Process.*, vol. 1, 2004, pp. I–225.

[6] T. Adali, H. Li, M. Novey, and J.-F. Cardoso, "Complex ICA using nonlinear functions," *IEEE Trans. Signal Process.*, vol. 56, no. 9, pp. 4536–4544, Sep. 2008.

[7] E. Bingham and A. Hyvärinen, "A fast fixed-point algorithm for independent component analysis of complex valued signals," *Int. J. Neural Syst.*, vol. 10, no. 01, pp. 1–8, Feb. 2000.

[8] M. Novey and T. Adali, "On extending the complex FastICA algorithm to noncircular sources," *IEEE Trans. Signal Process.*, vol. 56, no. 5, pp. 2148–2154, May 2008.

[9] ——, "Complex ICA by negentropy maximization," *IEEE Trans. Neural Networks,*, vol. 19, no. 4, pp. 596–609, Apr. 2008.

[10] X.-L. Li and T. Adali, "Complex independent component analysis by entropy bound minimization," *IEEE Trans. Circuits Syst. I Regul. Pap.*, vol. 57, no. 7, pp. 1417–1430, Jul. 2010.

[11] G.-S. Fu, R. Phlypo, M. Anderson, and T. Adali, "Complex independent component analysis using three types of diversity: Non-Gaussianity, nonwhiteness, and noncircularity," *IEEE Trans. Signal Process.*, vol. 63, no. 3, pp. 794–805, Feb. 2015.

[12] A. Hyvärinen, "Fast and robust fixed-point algorithms for independent component analysis," *IEEE Trans. Neural Networks,*, vol. 10, no. 3, pp. 626–634, May 1999.

[13] H. Chen, D. Yao, and L. Zeng, "A BFGS-ICA algorithm and application in localization of brain activities," *Neurocomputing*, vol. 64, pp. 513–519, 2005.

[14] P. Ablin, J. Cardoso, and A. Gramfort, "Faster Independent Component Analysis by Preconditioning With Hessian Approximations," *IEEE Trans. Signal Process.*, vol. 66, no. 15, pp. 4040–4049, Aug. 2018.

[15] P. Ablin, J.-F. Cardoso, and A. Gramfort, "Faster ICA under orthogonal constraint," *arXiv Prepr. arXiv1711.10873*, 2017.

[16] D. C. Liu and J. Nocedal, "On the limited memory BFGS method for large scale optimization," *Math. Program.*, vol. 45, no. 1-3, pp. 503–528, Aug. 1989.

[17] L. Sorber, M. V. Barel, and L. D. Lathauwer, "Unconstrained optimization of real functions in complex variables," *SIAM J. Optim.*, vol. 22, no. 3, pp. 879–898, Jul. 2012.

[18] T. Adali, P. J. Schreier, and L. L. Scharf, "Complex-Valued Signal Processing: The Proper Way to Deal With Impropriety," *IEEE Trans. Signal Process.*, vol. 59, no. 11, pp. 5101–5125, Nov. 2011.

[19] T. Adali and P. J. Schreier, "Optimization and estimation of complex-valued signals: Theory and applications in filtering and blind source separation," *IEEE Signal Process. Mag.*, vol. 5, no. 31, pp. 112–128, Sep. 2014.

[20] X.-L. Li and X.-D. Zhang, "Nonorthogonal joint diagonalization free of degenerate solution," *IEEE Trans. Signal Process.*, vol. 55, no. 5, pp. 1803–1814, May 2007.

[21] A. H. T. Nguyen, V. G. Reju, A. W. H. Khong, and Y. Soon, "Learning complex-valued latent filters with absolute cosine similarity," in *Proc. IEEE Int. Conf. Acoust. Speech, Signal Process.*, 2017, pp. 2412–2416.

[22] M. Schmidt, "MinFunc: Unconstrained differentiable multivariate optimization in Matlab," 2005. [Online]. Available: http://www.cs.ubc.ca/ schmidtm/Software/minFunc.html

[23] S. Amari, A. Cichocki, and H. H. Yang, "A New Learning Algorithm for Blind Signal Separation," *Adv. Neural Inf. Process. Syst.*, pp. 757–763, 1996.

[24] M. Novey, T. Adali, and A. Roy, "A complex generalized Gaussian distribution - Characterization, generation, and estimation," *IEEE Trans. Signal Process.*, vol. 58, no. 3, pp. 1427–1433, Mar. 2010.