A novel fixed-point algorithm for constrained independent component analysis

Guobing Qian[1,2], Lidan Wang[1*], Shiyuan Wang[1] and Shukai Duan[1]

Abstract

Constrained independent component analysis (ICA) is an effective method for solving the blind source separation with *a prior* knowledge. However, most constrained ICA algorithms are proposed for the real-valued sources. In this paper, a novel constrained noncircular complex fast independent component analysis (c-ncFastICA) algorithm based on the fixed-point learning is proposed to address the complex-valued sources. The c-ncFastICA algorithm uses the augmented Lagrangian method to obtain a new cost function and then utilizes the quasi-Newton method to search its optimal solution. Compared with other ICA and constrained ICA algorithms, c-ncFastICA has better separation performance. Simulations confirm the effectiveness and superiority of the c-ncFastICA algorithm.

Keywords: Constrained ICA, Noncircular, Complex, Fixed-point

1 Introduction

Independent component analysis (ICA) [1–3] is very effective in separating the linear mixture of independent sources. It has been widely used in many scenarios, such as in array signal processing [4], speech signal processing [5], and multivariate probability density function estimation [6]. During the past decades, a lot of ICA algorithms have been proposed for the separation of real-valued and complex-valued signals. These algorithms can be divided into the algebraic theory-based ones [7, 8], the information criterion-based ones [9–12], and the iterative projection-based ones [13, 14]. For example, Joint Approximate Diagonalization of Eigenmatrices (JADE) [7] and alternating columns-diagonal centers (AC-DC) [8] are algebraic theory-based algorithms; FastICA [9], complex FastICA (cFastICA) [10], and noncircular complex FastICA (ncFastICA) [11] are information criterion-based algorithms; and auxiliary function-based ICA [13] is the iterative projection-based algorithm. In addition, independent vector analysis (IVA) [15, 16] and independent low-rank matrix analysis (ILRMA) [17] are the extensions of ICA to multidimensional cases. In many practical applications, such as wind measurements and digital communication, the sources are noncircular complex-valued [18,

19]. Thus, it is necessary to develop an efficient noncircular complex ICA algorithm. Among the aforementioned ICA algorithms, one of the most effective and commonly used algorithms for noncircular complex sources is the ncFastICA algorithm [11]. It incorporates the noncircular information of the original sources into the fixed-point iteration and provides better separation performance under the case of noncircular sources. However, ncFastICA only exploits the independence and noncircular property of the original sources. It does not utilize the prior knowledge to improve the separation performance.

To overcome the defect of the ICA, constrained ICA has been proposed in [20–22]. It combines the prior information and the statistical independence of the signals and thus achieves better separation performance than ICA. The constrained ICA algorithms are mainly derived from information criterion, e.g., kurtosis-based constrained ICA algorithm [23], and maximum likelihood (ML)-based constrained ICA algorithm [24]. [23] shows better performance in extracting the sources of the mechanical system. [24] extends the solution space and thus has better performance than other constrained algorithms. However, both algorithms need to choose the threshold parameter. To overcome this defect, Shi et al. proposed a new constrained ICA algorithm based on multi-objective optimization recently [25]. The fixed-point learning is derived based on Kuhn-Tucker conditions, and it does not need to choose the threshold parameter. However, these

* Correspondence: ldwang@swu.edu.cn
[1]Chongqing Key Laboratory of Nonlinear Circuits and Intelligent Information Processing, College of Electronic and Information Engineering, Southwest University, Chongqing 400715, China
Full list of author information is available at the end of the article

algorithms are suitable only for the real-valued signals, which severely limit their applications. To extend the application of constrained ICA to complex sources, Wang et al. [26] proposed a class of complex constrained ICA algorithms based on gradient descent method, in which circular complex signals are considered. However, the aforementioned constrained ICA algorithms are not suitable for the general complex-valued signals.

In this paper, to improve the generality ability of the constrained complex ICA, we propose a new *constrained noncircular complex fast independent component analysis* (c-ncFastICA) algorithm. In c-ncFastICA, a new cost function is built using the augmented Lagrangian method. The prior information is then combined into the fixed-point iteration based on a quasi-Newton method. Stability analysis shows that the optimal solution corresponds to the fixed point of c-ncFastICA.

The rest of this paper is organized as follows: Section 2 presents the system model, and Section 3 derives the c-ncFastICA algorithm and gives the stability analysis. The superiority of the c-ncFastICA algorithm is verified by simulations in Sections 4, and the conclusion is drawn in Section 5.

2 System model
In this section, the complex ICA model and its constrained model are introduced respectively.

2.1 Complex ICA model
The basic complex ICA model is [10].

$$\mathbf{z}=\mathbf{As} \tag{1}$$

where $\mathbf{z} = [z_1 \; \cdots \; z_M]^T \in \mathbb{C}^{M \times 1}$ and $\mathbf{s} = [s_1 \; \cdots \; s_N]^T \in \mathbb{C}^{N \times 1}$ denote M observation signals and N original sources, respectively; $\mathbf{A} \in \mathbb{C}^{M \times N}$ is the mixing matrix; and notation $\mathbb{C}$ denotes the complex domain.

It is noted that the sources and mixing matrix are complex-valued in complex ICA model, which is the main difference between the complex ICA model and real ICA model. In general, it assumes that the original sources are zero mean and unit variance. In addition, $\mathbf{A}$ is of full-column rank, and at most one original source is Gaussian [27]. The purpose of complex ICA is to find a complex-valued demixing matrix $\overline{\mathbf{w}}$ for $\mathbf{z}$ to recover the original signal, with $\hat{\mathbf{s}} = \overline{\mathbf{w}}^H \mathbf{z}$ denoting the recovered signal.

2.2 Constrained complex ICA model
The constrained complex ICA model is described by the complex ICA model with following constraints

$$h_n(\overline{\mathbf{w}}_n, \mathbf{r}_n) = \rho_n - \varepsilon(\overline{\mathbf{w}}_n, \mathbf{r}_n) \leq 0$$
$$\text{and} \; f_n(\overline{\mathbf{w}}_n) = 0 \tag{2}$$

where h_n and f_n are the inequality constriction and the

equality constriction, respectively; ε and ρ_n are the measurement function and the threshold for the inequality constriction; $\overline{\mathbf{w}}_n$ is the nth column of the demixing matrix $\overline{\mathbf{w}}$; $\mathbf{r}_n$ is a reference vector for $\overline{\mathbf{w}}_n$.

Typically, ε is a distance measurement and can be described by an inner product in practice [24]. Thus, it is chosen as $\varepsilon(\overline{\mathbf{w}}_n, \mathbf{r}_n) = \left|\overline{\mathbf{w}}_n^H \mathbf{r}_n\right|^2$ in this paper.

3 Constrained ncFastICA algorithm
In this section, we incorporate the constraints into the original ncFastICA algorithm and thus obtain a new algorithm, namely constrained noncircular complex FastICA (c-ncFastICA) algorithm. The stability analysis shows that the optimal solution corresponds to the fixed point of the new algorithm.

3.1 Whitening
Whitening is necessary for the c-ncFastICA algorithm, which is to make the observation data $\mathbf{z}$ uncorrelated and unit variance. It is implemented by the transform

$$\mathbf{x}=\mathbf{Vz}, \quad \text{with} \quad \mathbf{V}=\boldsymbol{\Lambda}^{-1/2}\mathbf{U}^H \tag{3}$$

where $\mathbf{V}$ is the whitening matrix; $\mathbf{x}$ is the whitened data; $\boldsymbol{\Lambda}$ is a diagonal matrix, with N largest eigenvalues of $\mathbf{R} = E(\mathbf{z}\,\mathbf{z}^H)$ as its entries; $\mathbf{U}$ is a matrix consisting of the corresponding eigenvectors; and N is the source number.

After whitening, the demixing matrix $\mathbf{w}$ for $\mathbf{x}$ is a unitary matrix and the recovered signal can be represented by $\hat{\mathbf{s}} = \mathbf{w}^H \mathbf{x}$. The relationship between $\mathbf{w}$ and $\overline{\mathbf{w}}$ is $\overline{\mathbf{w}}^H = \mathbf{w}^H \mathbf{V}$.

3.2 Cost function
As shown in [11], the cost function of original ncFastICA algorithm is

$$J(\mathbf{w}_n)=E\left\{G\left(\left|\mathbf{w}_n^H\mathbf{x}\right|^2\right)\right\} \tag{4}$$

where $G: \mathbb{R}^+ \cup \{0\} \to \mathbb{R}$ is a smooth function, $\mathbb{R}$ and $\mathbb{R}^+$ correspond to the real domain and positive real domain, respectively; $\mathbf{w}_n \in \mathbb{C}^N$ is the nth column of the demixing matrix $\mathbf{w}$, with $\|\mathbf{w}\|_2 = 1$.

In practice, there are three choices for G [11]: $G_1(u) = \log(0.1 + u)$, $G_2(u) = \sqrt{0.1 + u}$, $G_3(u) = \frac{1}{2}u^2$. G_1 and G_2 provide more robust estimators, and G_3 is motivated by kurtosis. By using the augmented Lagrangian method [24], the inequality constraint in (2) can be incorporated into the cost function.

Then, the new cost function can be derived as

$$J^c(\mathbf{w}_n, \mu_n)=E\left\{G\left(\left|\mathbf{w}_n^H\mathbf{x}\right|^2\right)\right\}$$
$$+ \frac{1}{2\gamma_n}\left(\left(\max\{0, \gamma_n h_n(\mathbf{w}_n, \mathbf{r}_n) + \mu_n\}\right)^2 - \mu_n^2\right) \tag{5}$$

where μ^n and γ^n are the Lagrangian multiplier and positive learning parameter, respectively, with $\|\mathbf{w}^n\|^2 = 1$. Thus, the equality constriction is expressed as $f_n(\mathbf{w}_n) = \mathbf{w}_n \mathbf{w}_n^H - 1 = 0$.

3.3 Fixed-point iteration

By using the quasi-Newton methods, we derive the one unit iteration of c-ncFastICA as

$$\mu_n \leftarrow \max\{\gamma_n h_n(\mathbf{w}_n, \mathbf{r}_n) + \mu_n, 0\} \tag{6}$$

$$
\begin{aligned}
\mathbf{w}_n^{(m+1)} = &-E\{g(|y|^2)y^*\mathbf{x}\} + E\{g'(|y|^2)|y|^2 + g(|y|^2)\}\mathbf{w}_n^{(m)} \\
&+ E\{\mathbf{x}\mathbf{x}^T\}E\{g'(|y|^2)y^{*2}\}\mathbf{w}_n^{(m)*} \\
&+ \operatorname{sign}(\mu_n) \cdot 2\gamma_n |\mathbf{w}_n^{(m)H}\mathbf{r}_n|^2 (\mathbf{w}_n^{(m)H}\mathbf{r}_n)^* \mathbf{r}_n
\end{aligned}
\tag{7}
$$

where $y = \mathbf{w}_n^{(m)H}\mathbf{x}$, $g(u)$, and $g'(u)$ denote the first-order and second-order derivative of $G(u)$; m represents the number of iteration. The detailed derivation can be seen in Appendix A.

The one unit iteration can be extended to the whole demixing matrix $\mathbf{w}$ by the symmetric way or the deflation way [10]. In the symmetric way, $\mathbf{w}$ is orthogonalized by $\mathbf{w} \leftarrow (\mathbf{w}\mathbf{w}^H)^{1/2}\mathbf{w}$. In the deflation way, $\mathbf{w}$ is orthogonalized by a Gram-Schmidt-like method. In this paper, we use the symmetric scheme.

In short, the c-ncFastICA algorithm is summarized in Table 1.

3.4 Stability analysis

In this subsection, we present the stability analysis of the fixed-point iteration. We show that the optimal solution to the c-ncFastICA algorithm corresponds to the fixed point of the (7). The detailed proof is shown as follows:

Proof By making the orthogonal transformation $\mathbf{q}_n = (\mathbf{V}\mathbf{A})^H\mathbf{w}_n$, the fixed-point iteration for the constraint optimization problem (5) becomes

Table 1 c-ncFastICA algorithm

Step 1. Whiten the observed data to make it uncorrelated and unit variance;

Step 2. Initialize the demixing matrix as an identity matrix, $\mu_n = 0$, γ_n to be a small positive valued number, set ρ_n and G;

Step 3. If the reference is available, update $\mathbf{w}$ based on (7); otherwise, update $\mathbf{w}$ based on the fixed-point iteration of original ncFastICA;

Step 4. Orthogonalize the demixing matrix $\mathbf{W}$;

Step 5. Repeat steps 3 and 4 until convergence;

Step 6. Compute the demixing matrix $\overline{\mathbf{W}} = (\mathbf{W}^H \mathbf{V})^H$ and obtain the recovered signal $\hat{\mathbf{s}} = \overline{\mathbf{W}}^H \mathbf{z}$.

$$
\begin{aligned}
\mathbf{q}_n^{(m+1)} = &-E\{g(|y|^2)y^*\mathbf{s}\} + E\{g'(|y|^2)|y|^2 + g(|y|^2)\}\mathbf{q}_n^{(m)} \\
&+ E\{\mathbf{s}\mathbf{s}^T\}E\{g'(|y|^2)y^{*2}\}\mathbf{q}_n^{(m)*} \\
&+ \operatorname{sign}\left(\max\{\gamma_n h_n(\mathbf{q}_n^{(m)}, \bar{\mathbf{r}}_n) + \mu_n, 0\}\right) \\
&\cdot 2\gamma_n |\mathbf{q}_n^{(m)H}\bar{\mathbf{r}}_n|^2 (\mathbf{q}_n^{(m)H}\bar{\mathbf{r}}_n)^* \bar{\mathbf{r}}_n
\end{aligned}
\tag{8}
$$

where $\bar{\mathbf{r}}_n = (\mathbf{V}\mathbf{A})^H \mathbf{r}_n$.

As we know, the optimal solution $\mathbf{q}_n$ of c-ncFastICA algorithm satisfies the conditions that only the nth element is non-zero and $\|\mathbf{q}_n\|_2 = 1$. Without loss of generality, the non-zero element is assumed to be $e^{j\theta_n}$. Thus, $y = \mathbf{q}_n^{(m)H}\mathbf{s} = e^{-j\theta_n}s_n$ and

$$
\begin{aligned}
\mathbf{q}_n^{(m+1)} = &-E\{g(|s_n|^2)e^{j\theta_n}s_n^*\mathbf{s}\} + E\{g'(|s_n|^2)|s_n|^2 + g(|s_n|^2)\}\mathbf{q}_n^{(m)} \\
&+ E\{\mathbf{s}\mathbf{s}^T\}E\{g'(|s_n|^2)y^{*2}\}\mathbf{q}_n^{(m)*} \\
&+ \operatorname{sign}\left(\max\{\gamma_n h_n(\mathbf{q}_n^{(m)}, \bar{\mathbf{r}}_n) + \mu_n, 0\}\right) \\
&\cdot 2\gamma_n |\mathbf{q}_n^{(m)H}\bar{\mathbf{r}}_n|^2 (\mathbf{q}_n^{(m)H}\bar{\mathbf{r}}_n)^* \bar{\mathbf{r}}_n
\end{aligned}
\tag{9}
$$

For the first term in the right hand of Eq. (9), it can be derived that the nth element is $-E\{g(|s_n|^2)|s_n|^2\}e^{j\theta_n}$ and other elements are zero since the original sources are independent. Thus, the first term is rewritten as $-E\{g(|s_n|^2)|s_n|^2\}\mathbf{q}_n^{(m)}$. Similarly, for the third term in the right hand of Eq. (9), the nth element is $E\{s_n^2\}E\{g'(|s_n|^2)s_n^{*2}\}e^{j\theta_n}$ and other elements are zero. For the fourth term, the nth element is $\operatorname{sign}(\max\{\gamma_n h_n(\mathbf{q}_n^{(m)}, \bar{\mathbf{r}}_n) + \mu_n, 0\}) \cdot 2\gamma_n e^{j\theta_n}$ and other elements are zero.

Thus, (9) is rewritten as

$$\mathbf{q}_n^{(m+1)} = (\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4)\mathbf{q}_n^{(m)} \tag{10}$$

where $\alpha_1 = -E\{g(|s_n|^2)|s_n|^2\}$, $\alpha_2 = E\{g'(|s_n|^2)|s_n|^2 + g(|s_n|^2)\}$, $\alpha_3 = E\{s_n^2\}E\{g'(|s_n|^2)s_n^{*2}\}$, and $\alpha_4 = \operatorname{sign}(\max\{\gamma_n h_n(\mathbf{q}_n^{(m)}, \bar{\mathbf{r}}_n) + \mu_n, 0\}) \cdot 2\gamma_n$.

Considering the constraint of $\|\mathbf{q}_n\|_2 = \|\mathbf{w}_n\|_2 = 1$, we can remove the real-valued coefficient $\alpha = \alpha_1 + \alpha_2 + \alpha_3 + \alpha_4$ from (10) if it is not equal to zero. Therefore, $\mathbf{q}_n^{(m)}$ is the fixed point of (10).

Remark

1) The parameters γ_n, μ_n, and ρ_n influence the convergence of the c-ncFastICA algorithm. If $\gamma_n h_n(\mathbf{q}_n^{(m)}, \bar{\mathbf{r}}_n) + \mu_n \leq 0$, then $\alpha_4 = 0$. In this case, the constraints have no effect on the c-ncFastICA algorithm. Thus, the convergence of the

c-ncFastICA algorithm is the same as the original ncFastICA algorithm. On the contrary, if $\gamma_n h_n(\mathbf{q}_n^{(m)}, \bar{\mathbf{r}}_n) + \mu_n > 0$, then $\alpha_4 = 2\gamma_n$. In this case, the value of γ_n should be chosen to make sure the real-valued coefficient α is not equal or close to zero.

2) If the prior information is accurate enough, it is suggested to choose a big ρ_n; if the prior information is not accurate enough, it is suggested to choose a big ρ_n at the first several iterations and then choose a small ρ_n at the rest iterations. In practice, both G and ρ_n are usually chosen by trials.

3) As a byproduct, we can also see that the assumptions for the constrained sources can be a little relaxed in the constrained complex ICA model, i.e., the constrained source s_n can be Gaussian, since γ_n can be properly chosen to make sure the real-valued coefficient is not equal to zero even if the source is Gaussian ($-E\{g(|s_n|^2)|s_n|^2\} + E\{g'(|s_n|^2)|s_n|^2 + g(|s_n|^2)\} + E\{s_n^2\}E\{g'(|s_n|^2)s_n^{*2}\} = 0$).

4 Results and discussion

In this section, the superiority of the c-ncFastICA algorithm is demonstrated by some simulations. We employ the normalized Amari index [28] to measure the performance of the different algorithms

$$
I_A = \frac{1}{2M(M-1)}\left[\sum_{i=1}^{M}\left(\sum_{j=1}^{M}\frac{|p_{ij}|}{\max_k|p_{ik}|}-1\right)\right.
$$
$$
\left.\times + \sum_{j=1}^{M}\left(\sum_{i=1}^{M}\frac{|p_{ij}|}{\max_k|p_{kj}|}-1\right)\right] \qquad (11)
$$

where $\mathbf{P} = \mathbf{w}^H\mathbf{VA}$. The lower of I_A indicates the better of the separation. For all the simulations, the Amari index is obtained by the average of 100 Monte Carlo trials. We compare the c-ncFastICA algorithm with the symmetric gradient descent method [26] since it is the only algorithm which can be applied to complex-valued signals among the constrained ICA algorithms in [20–26]. In addition, we also compare c-ncFastICA with the ncFastICA algorithm since it is widely used for the separation of noncircular signals.

In the first simulation, we use the complex generalized Gaussian distributed (cGGD) signals [11] as the original sources. For each trial, the source number is eight and half of the sources are made noncircular with the same noncircularity index. The shape parameters and noncircularity index for each source are listed in Table 2. The sample size of the original sources varies from 250 to 1000. The complex-valued mixing matrix $\mathbf{A}$ is 8×8 dimensional.

Table 2 The shape parameters and noncircularity index for the original sources

Source	1	2	3	4
Parameter, p	0.9	0.7	0.9	0.7
Noncircularity index	1	1	1	1
Source	5	6	7	8
Parameter p	2	2	1.1	2
Noncircularity index	9.95	9.95	9.95	9.95

Each element is zero mean and unit variance, with the real part and imaginary part normally distributed. We use the third and seventh columns of $\mathbf{A}$ as reference.

Figure 1 shows the time sequences and histograms of original sources 3 and 7, and Fig. 2 shows the time sequences and histograms of recovered sources 3 and 7 by c-ncFastICA algorithm, where $\rho_n = 0.9$, $\gamma_n = 3$, and the nonlinear function is selected as G_3. One can see that the constrained sources 3 and 7 are successfully recovered by c-ncFastICA even though they are almost Gaussian. Figure 3 depicts the separation results of c-ncFastICA algorithm, ncFastICA algorithm, and the symmetric gradient descent method [26], where $\rho_n = 0.9$ and $\gamma_n = 3$. It shows that the c-ncFastICA algorithm performs significantly better than other two methods. This is due to the fact that ncFastICA does not take the constrained condition into consideration, and symmetric gradient descent method cannot make full use of the noncircular characteristic of the sources. Moreover, the performance gap widens slightly with the increase of the sample size. It also shows that the ncFastICA algorithm cannot perform well even when the sample size is 1000. This is due to the fact that source 1, source 3, and source 7 are almost Gaussian in this simulation. On the contrary, the c-ncFastICA algorithm can get desirable separation result in this case, which verifies the stability analysis in the Section 3. Figure 4 compares the convergence curves of c-ncFastICA under different nonlinear functions, with the sample size fixed at 1000. One can observe that c-ncFastICA converges quickly under different nonlinear functions. Figure 5 shows the influence of ρ_n on the performance of c-ncFastICA, with the sample size fixed at 1000. It can be seen that the algorithm performs better when ρ_n is bigger, which is due to the fact that the prior information is accurate in this simulation.

In the second simulation, we use three real-world frequency-modulated (FM) signals as the original sources. The powers of three sources are 1, 1, and 10, respectively. The first and third sources are the interference signals, and the second source is the desired signal. For each trial, the parameters for carrier frequency and maximum frequency deviation are set as 80 kHz and ± 75 kHz, respectively. The original FM sources are received by the linear uniform
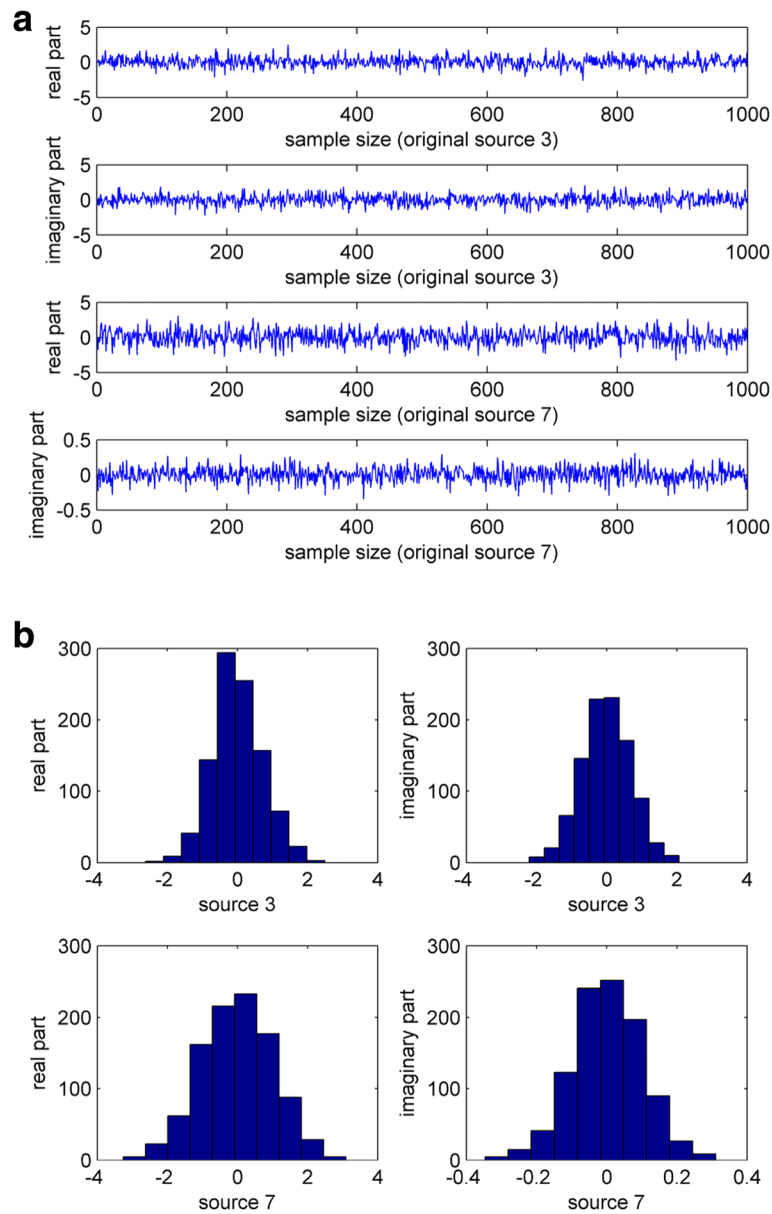
Fig. 1 a Time sequences of original sources 3 and 7. **b** Histograms of original sources 3 and 7

array, and the sensor number is five. The array interval is equal to the half of the wavelength. In this case, the mixing matrix is related to the directions of arrival (DOAs) of the original sources. The DOAs of the original sources are 30°, − 8°, and − 7°, respectively. We assume that the DOA of desired signal can be roughly detected, and thus, the second column of A (denoted as a_2) is considered as a prior knowledge. The additive white Gaussian noise is added to the receiver and the variance is 0.1. Figure 6 depicts the separation results of the c-ncFastICA algorithm, ncFastICA algorithm, and symmetric gradient descent method [26], where $\rho_n = 0.95$ and $\gamma_n = 3$. One can see that the

c-ncFastICA algorithm performs better than other two methods. Figure 7 shows the influence of ρ_n on the performance of c-ncFastICA algorithm. It can be seen that the performance of c-ncFastICA improves a lot when $\rho_n \geq 0.85$ (G_1 and G_3) or $\rho_n \geq 0.9$ (G_2). However, this does not mean that the bigger the ρ_n, the better the performance. Figure 8 shows the influence of the accuracy of prior knowledge on the performance of c-ncFastICA, where $\rho_n = 0.95$, $\gamma_n = 3$, and the prior knowledge for each element of a_2 is contaminated by Gaussian noise with zero mean and σ^2. One can see that with $\rho_n = 0.95$, the performance of c-ncFastICA becomes worse when the prior knowledge is not accurate.
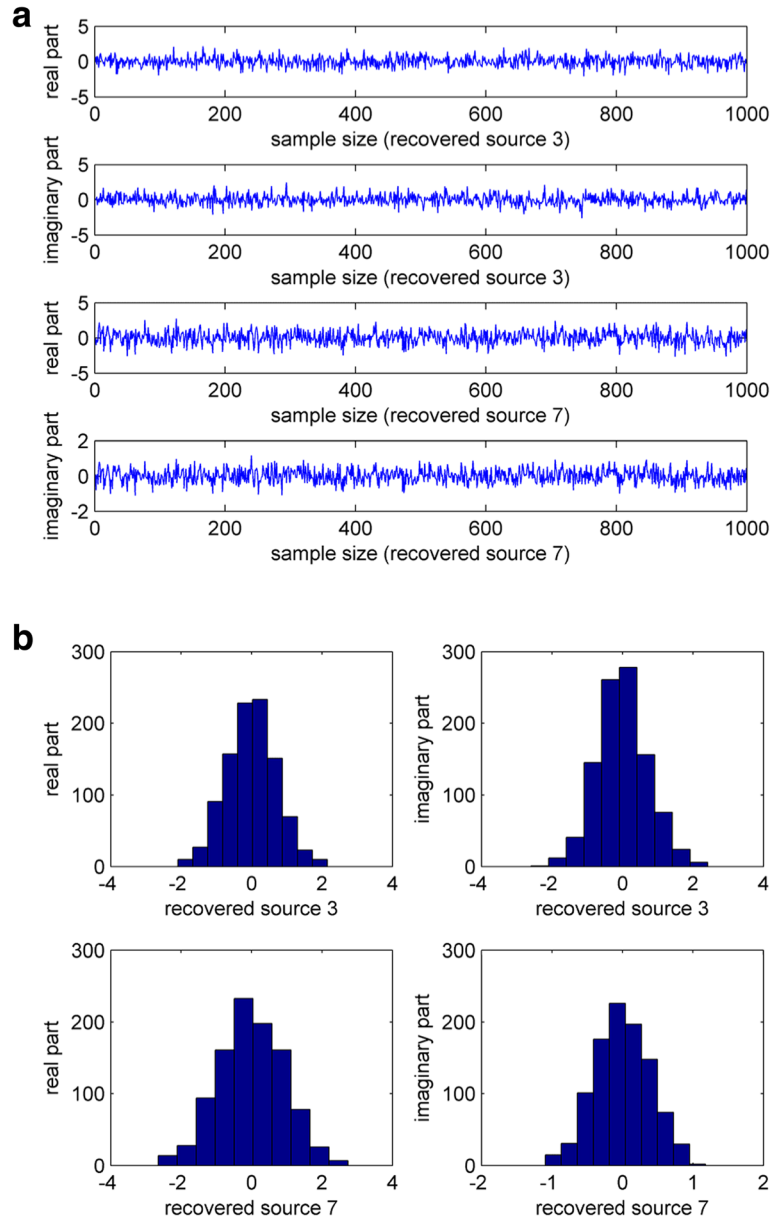
Fig. 2 a Time sequences of recovered sources 3 and 7. **b** Histograms of recovered sources 3 and 7

5 Conclusion

The constrained ICA for complex sources is a challenging problem. In this paper, we focus on this problem and extend the noncircular complex FastICA (ncFastICA) algorithm to the constrained case. By adding the constrained conditions to the cost function and utilizing the quasi-Newton method, we derive a new fixed-point algorithm, namely constrained ncFastICA (c-ncFastICA) algorithm. Stability analysis shows that the optimal solution to constrained ICA corresponds to the fixed point of the c-ncFastICA algorithm. Simulations verify the correctness of the stability analysis and the superiority of the c-ncFastICA algorithm.

6 Appendix

6.1 Derivation of the fixed-point algorithm

The Lagrangian function for the constrained ICA is

$$L^c(\mathbf{w}_n, \lambda, \mu_n) = J^c(\mathbf{w}_n) + \lambda\left(\mathbf{w}_n^H \mathbf{w}_n - 1\right) \tag{12}$$

As the cost function is not analytic in $\mathbf{w}_n$ but analytic in $\mathbf{w}_n$ and $\mathbf{w}_n^*$ independently [11], we derive the quasi-Newton update as follows:

Fig. 3 Separation results for cGGD signals

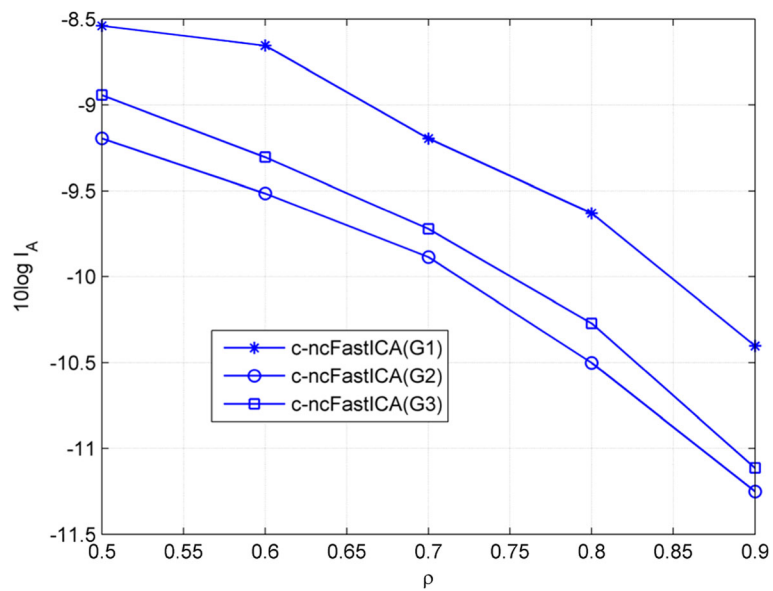**Fig. 4** Convergence curves of c-ncFastICA

Fig. 5 Influence of ρ_n on the performance of c-ncFastICA
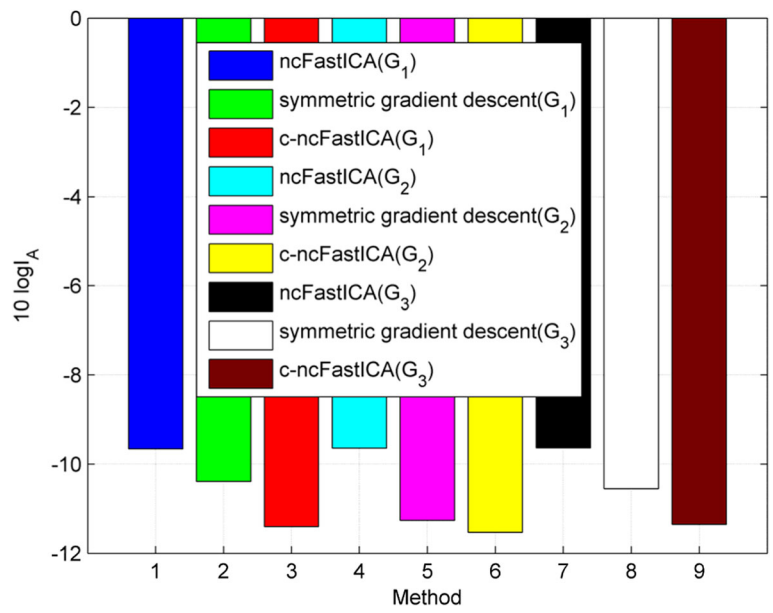

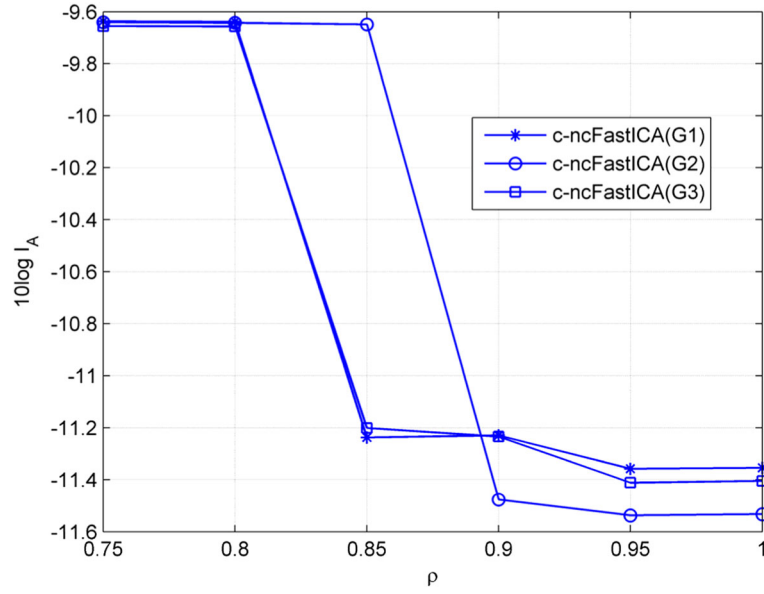**Fig. 6** Separation results of different methods

Fig. 7 Influence of ρ_n on the performance of c-ncFastICA

$$\Delta\tilde{\mathbf{w}}_n = -\left(\frac{\partial^2 L^c}{\partial\tilde{\mathbf{w}}_n^* \partial\tilde{\mathbf{w}}_n^T}\bigg|_{\mathbf{w}_n=\mathbf{w}_n^{(m)}}\right)^{-1}\frac{\partial L^c}{\partial\tilde{\mathbf{w}}_n^*}\bigg|_{\mathbf{w}_n=\mathbf{w}_n^{(m)}} \quad (13)$$

$$= -\left(\tilde{\mathbf{H}}_{\mathbf{w}_n}L^c\right)^{-1}\tilde{\nabla}_{\mathbf{w}_n}^* L^c$$

where

$$\begin{aligned}\mathbf{w}_n &= [w_1, w_2, ..., w_N]^T \in \mathbb{C}^N \\ \tilde{\mathbf{w}}_n &= [w_1, w_1^*, ..., w_N, w_N^*]^T \in \mathbb{C}^{2N}\end{aligned} \quad (14)$$

and $\Delta\tilde{\mathbf{w}}_n = \tilde{\mathbf{w}}_n^{(m+1)} - \tilde{\mathbf{w}}_n^{(m)}$; $\tilde{\mathbf{H}}_{\mathbf{w}_n}L^c$ and $\tilde{\nabla}_{\mathbf{w}_n}L^c$ are the

Hessian matrix and gradient vector of the Lagrangian function, respectively.

Combining (12, 13), we can derive

$$\Delta\tilde{\mathbf{w}}_n = -\left(\tilde{\mathbf{H}}_{\mathbf{w}_n}J^c + \lambda\tilde{\mathbf{I}}\right)^{-1}\left(\tilde{\nabla}_{\mathbf{w}_n}^* J^c + \lambda\tilde{\mathbf{w}}_n^{(m)}\right) \quad (15)$$

and

$$\left(\tilde{\mathbf{H}}_{\mathbf{w}_n}J^c + \lambda\tilde{\mathbf{I}}\right)\tilde{\mathbf{w}}_n^{(m+1)} = -\tilde{\nabla}_{\mathbf{w}_n}^* J^c + \tilde{\mathbf{H}}_{\mathbf{w}_n}J^c\tilde{\mathbf{w}}_n^{(m)} \quad (16)$$

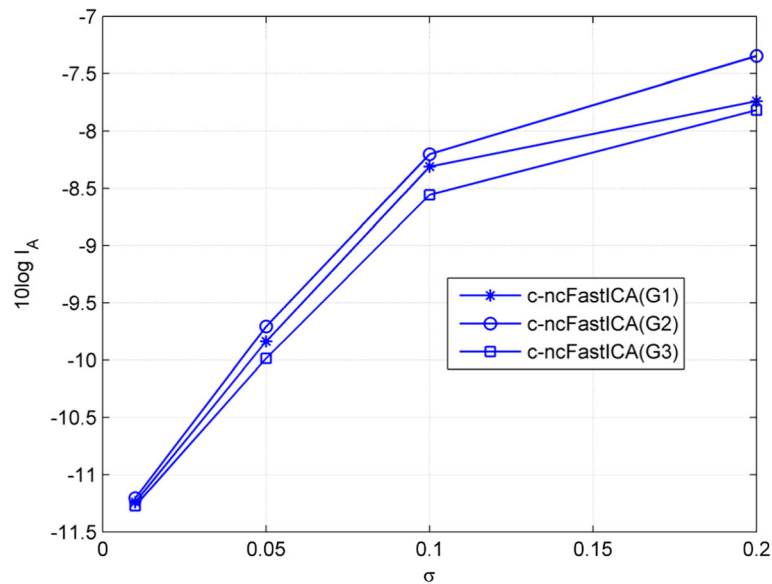1) If $\gamma_n h_n(\mathbf{w}_n, \mathbf{r}_n) + \mu_n > 0$,

Fig. 8 Influence of the accuracy of prior knowledge

$$\tilde{\nabla}_{\mathbf{w}_n} J^c = E \begin{pmatrix} \frac{\partial J^c}{\partial w_1} \\ \frac{\partial J^c}{\partial w_1^*} \\ \vdots \\ \frac{\partial J^c}{\partial w_N} \\ \frac{\partial J^c}{\partial w_N^*} \end{pmatrix} = \begin{pmatrix} E\{g(yy^*)yx_1^*\} \\ E\{g(yy^*)y^*x_1\} \\ \vdots \\ E\{g(yy^*)yx_N^*\} \\ E\{g(yy^*)y^*x_N\} \end{pmatrix} - (\gamma_n h_n + \mu_n) \begin{pmatrix} \left(\mathbf{w}_n^{(m)H}\mathbf{r}_n\right)\mathbf{r}_1^* \\ \left(\mathbf{w}_n^{(m)H}\mathbf{r}_n\right)^*\mathbf{r}_1 \\ \vdots \\ \left(\mathbf{w}_n^{(m)H}\mathbf{r}_n\right)\mathbf{r}_N^* \\ \left(\mathbf{w}_n^{(m)H}\mathbf{r}_n\right)^*\mathbf{r}_N \end{pmatrix} \tag{17}$$

$$\tilde{\mathbf{H}}_{\mathbf{w}_n} J^c = E\left(\frac{\partial^2 J^c}{\partial \tilde{\mathbf{w}}_n^* \partial \tilde{\mathbf{w}}_n^T}\right) = E\left\{ \begin{bmatrix} x_1 x_1^* e & x_1^2 d & \cdots & x_1 x_N^* e & x_1 x_N d \\ x_1^{*2} d^* & \hat{x}_1^* \hat{x}_1 e & \cdots & x_1^* x_N^* d^* & x_1^* x_N e \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ x_N^* x_1^* d^* & x_N^* x_1 e & \cdots & x_N^* x_N^* d^* & x_N^* x_N e \end{bmatrix} \right\}$$

$$- (\gamma_n h_n + \mu_n) \begin{bmatrix} \left(\mathbf{w}_n^{(m)H}\mathbf{r}_n\right)^* \mathbf{r}_1 \mathbf{r}_1^* & 0 & \cdots & \left(\mathbf{w}_n^{(m)H}\mathbf{r}_n\right)^* \mathbf{r}_1 \mathbf{r}_N^* & 0 \\ 0 & \left(\mathbf{w}_n^{(m)H}\mathbf{r}_n\right)\mathbf{r}_1^* \mathbf{r}_1 & \cdots & 0 & \left(\mathbf{w}_n^{(m)H}\mathbf{r}_n\right)\mathbf{r}_1^* \mathbf{r}_N \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & \left(\mathbf{w}_n^{(m)H}\mathbf{r}_n\right)\mathbf{r}_N^* \mathbf{r}_1 & \cdots & & \left(\mathbf{w}_n^{(m)H}\mathbf{r}_n\right)\mathbf{r}_N^* \mathbf{r}_N \end{bmatrix}$$

$$+ \begin{pmatrix} \left(\mathbf{w}_n^{(m)H}\mathbf{r}_n\right)^* \mathbf{r}_1 \\ \left(\mathbf{w}_n^{(m)H}\mathbf{r}_n\right)\mathbf{r}_1^* \\ \vdots \\ \left(\mathbf{w}_n^{(m)H}\mathbf{r}_n\right)^* \mathbf{r}_N \\ \left(\mathbf{w}_n^{(m)H}\mathbf{r}_n\right)\mathbf{r}_N^* \end{pmatrix} \begin{bmatrix} \left(\mathbf{w}_n^{(m)H}\mathbf{r}_n\right)\mathbf{r}_1^* & \left(\mathbf{w}_n^{(m)H}\mathbf{r}_n\right)^* \mathbf{r}_1 & \cdots & \left(\mathbf{w}_n^{(m)H}\mathbf{r}_n\right)\mathbf{r}_N^* & \left(\mathbf{w}_n^{(m)H}\mathbf{r}_n\right)^* \mathbf{r}_N \end{bmatrix} \tag{18}$$

where $e = g'(|y|^2)|y|^2 + g(|y|^2)$, $d = g'(|y|^2)y^{*2}$.

Rewrite the $\tilde{\mathbf{H}}_{\mathbf{w}_n} J^c$ as two parts

$$\tilde{\mathbf{H}}_{\mathbf{w}_n} J^c = \tilde{\mathbf{H}}_{\mathbf{w}_n}^a J^c + \tilde{\mathbf{H}}_{\mathbf{w}_n}^b J^c \tag{19}$$

where

$$\tilde{\mathbf{H}}_{\mathbf{w}_n}^a J^c = E\left\{ \begin{bmatrix} x_1 x_1^* e & 0 & \cdots & x_1 x_N^* e & 0 \\ 0 & x_1^* x_1 e & \cdots & 0 & x_1^* x_N e \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & x_N^* x_1 e & \cdots & 0 & x_N^* x_N e \end{bmatrix} \right\} - (\gamma_n h_n + \mu_n) \begin{bmatrix} \mathbf{r}_1 \mathbf{r}_1^* & 0 & \cdots & \mathbf{r}_1 \mathbf{r}_N^* & 0 \\ 0 & \mathbf{r}_1^* \mathbf{r}_1 & \cdots & 0 & \mathbf{r}_1^* \mathbf{r}_N \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & \mathbf{r}_N^* \mathbf{r}_1 & \cdots & & \mathbf{r}_N^* \mathbf{r}_N \end{bmatrix}$$

$$+ \gamma_n \begin{bmatrix} \left|\mathbf{w}_n^{(m)H}\mathbf{r}_n\right|^2 \mathbf{r}_1 \mathbf{r}_1^* & 0 & \cdots & \left|\mathbf{w}_n^{(m)H}\mathbf{r}_n\right|^2 \mathbf{r}_1 \mathbf{r}_N^* & 0 \\ 0 & \left(\mathbf{w}_n^{(m)H}\mathbf{r}_n\right)^2 \mathbf{r}_1^* \mathbf{r}_1 & \cdots & 0 & \left(\mathbf{w}_n^{(m)H}\mathbf{r}_n\right)^2 \mathbf{r}_1^* \mathbf{r}_N \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & \left(\mathbf{w}_n^{(m)H}\mathbf{r}_n\right)^2 \mathbf{r}_N^* \mathbf{r}_1 & \cdots & 0 & \left(\mathbf{w}_n^{(m)H}\mathbf{r}_n\right)^2 \mathbf{r}_N^* \mathbf{r}_N \end{bmatrix} \tag{20}$$

and

Then,

$$\mathbf{H}_{\mathbf{w}_n}^b J^c = E\left\{ \begin{bmatrix} 0 & x_1^2 d & \cdots & 0 & x_1 x_N d \\ x_1^{*2} d^* & 0 & \cdots & x_1^* x_N^* d^* & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ x_N^* x_1^* d^* & 0 & \cdots & x_N^* x_N^* d^* & 0 \end{bmatrix} \right\} + \gamma_n \begin{bmatrix} 0 & \left(\mathbf{w}_n^{(m)H}\mathbf{r}_n\right)^{*2} \mathbf{r}_1^2 & \cdots & 0 & \left(\mathbf{w}_n^{(m)H}\mathbf{r}_n\right)^{*2} \mathbf{r}_1 \mathbf{r}_N \\ \left(\mathbf{w}_n^{(m)H}\mathbf{r}_n\right)^2 \mathbf{r}_1^{*2} & 0 & \cdots & \left(\mathbf{w}_n^{(m)H}\mathbf{r}_n\right)^2 \mathbf{r}_1^* \mathbf{r}_N^* & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \left(\mathbf{w}_n^{(m)H}\mathbf{r}_n\right)^2 \mathbf{r}_N^* \mathbf{r}_1^* & 0 & \cdots & \left(\mathbf{w}_n^{(m)H}\mathbf{r}_n\right)^2 \mathbf{r}_N^{*2} & 0 \end{bmatrix} \tag{21}$$

$$\tilde{\mathbf{H}}_{\mathbf{w}_n} J\tilde{\mathbf{w}}_n = \tilde{\mathbf{H}}_{\mathbf{w}_n}^a J\tilde{\mathbf{w}}_n + \tilde{\mathbf{H}}_{\mathbf{w}_n}^b J\tilde{\mathbf{w}}_n \tag{22}$$

Taking (20) and (21) into consideration, and retaining only the odd-numbered rows, we can obtain

$$\begin{aligned}
\mathbf{H}_{\mathbf{w}_n} J\mathbf{w}_n &= E\Big\{g'\big(|y|^2\big)|y|^2 + g\big(|y|^2\big)\Big\}\mathbf{w}_n^{(m)} \\
&+ E\{\mathbf{x}\mathbf{x}^T\}E\Big\{g'\big(|y|^2\big)y^{*2}\Big\}\mathbf{w}_n^{(m)*} \\
&+ 2\gamma_n\big|\mathbf{w}_n^{(m)H}\mathbf{r}_n\big|^2\big(\mathbf{w}_n^{(m)H}\mathbf{r}_n\big)^*\mathbf{r}_n \\
&- \big(\gamma_n h_n + \mu_n\big)\mathbf{r}_n\mathbf{r}_n^H\mathbf{w}_n^{(m)}
\end{aligned} \tag{23}$$

Similarly, we can get

$$\mathbf{K}\mathbf{w}_n^{(m+1)} = -\nabla_{\mathbf{w}_n}^* J + \mathbf{H}_{\mathbf{w}_n} J\mathbf{w}_n^{(m)} \tag{24}$$

where $\mathbf{K} = (\mathbf{H}J^c + \lambda\mathbf{I})$.

Combining (23) and (24), we can obtain

$$\begin{aligned}
\mathbf{K}\mathbf{w}_n^{(m+1)} &= -E\{g(|y|^2)y^*\mathbf{x}\} + E\Big\{g'\big(|y|^2\big)|y|^2 + g\big(|y|^2\big)\Big\}\mathbf{w}_n^{(m)} \\
&+ E\{\mathbf{x}\mathbf{x}^T\}E\Big\{g'\big(|y|^2\big)y^{*2}\Big\}\mathbf{w}_n^{(m)*} \\
&+ 2\gamma_n\big|\mathbf{w}_n^{(m)H}\mathbf{r}_n\big|^2\big(\mathbf{w}_n^{(m)H}\mathbf{r}_n\big)^*\mathbf{r}_n
\end{aligned} \tag{25}$$

At the optimal solution,

$$\begin{aligned}
\{\overline{\mathbf{K}}\mathbf{q}_n\}_n &= \big(2E\big\{g'\big(|s_n|^2\big)|s_n|^4\big\} + E\big\{g\big(|s_n|^2\big)|s_n|^2\big\} \\
&+ \lambda + 2\gamma_n - (\gamma_n h_n + \mu_n)\big)e^{j\theta_n}
\end{aligned} \tag{26}$$

and $\{\overline{\mathbf{K}}\mathbf{q}_n\}_i = 0 \quad i \neq n$

where $\mathbf{q}_n = (\mathbf{V}\mathbf{A})^H\mathbf{w}_n$ and $\overline{\mathbf{K}} = \mathbf{K}(\mathbf{V}\mathbf{A})$.

The Lagrangian function for $\mathbf{q}_n, \lambda, \mu_n$ can be written as

$$L^c(\mathbf{q}_n, \lambda, \mu_n) = J^c(\mathbf{q}_n) + \lambda\big(\mathbf{q}_n^H\mathbf{q}_n - 1\big). \tag{27}$$

Calculating the derivative of Lagrangian function for $\mathbf{q}_n$ at the optimal solution

$$\begin{aligned}
\frac{\partial L^c(\mathbf{q}_n, \lambda, \mu_n)}{\partial \mathbf{q}_n^*} &= E\big\{g\big(|s_i|^2\big)|s_i|^2\big\}\mathbf{q}_n \\
&+ \lambda\mathbf{q}_n - (\gamma_n h_n + \mu_n)\mathbf{q}_n
\end{aligned} \tag{28}$$

and solving $\frac{\partial L^c(\mathbf{q}_n, \lambda, \mu_n)}{\partial \mathbf{q}_n^*} = \mathbf{0}$, we can get

$$\lambda = -E\big\{g\big(|s_i|^2\big)|s_i|^2\big\} + \big(\gamma_n h_n + \mu_n\big) \tag{29}$$

Thus,

$$\overline{\mathbf{K}}\mathbf{q}_n = k\mathbf{q}_n \tag{30}$$

i.e.,

$$\mathbf{K}\mathbf{w}_n = k\mathbf{w}_n \tag{31}$$

where

$$k = 2E\big\{g'\big(|s_n|^2\big)|s_n|^4\big\} + 2\gamma_n \in \mathbb{R} \tag{32}$$

Therefore, $\mathbf{K}$ can be removed due to the constraints $\|\mathbf{w}_n\| = 1$. The fixed-point iteration becomes

$$\begin{aligned}
\mathbf{w}^{(m+1)} &= -E\{g(|y|^2)y^*\mathbf{x}\} + E\Big\{g'\big(|y|^2\big)|y|^2 + g\big(|y|^2\big)\Big\}\mathbf{w}_n^{(m)} \\
&+ E\{\mathbf{x}\mathbf{x}^T\}E\Big\{g'\big(|y|^2\big)y^{*2}\Big\}\mathbf{w}_n^{(m)*} \\
&+ 2\gamma_n\big|\mathbf{w}_n^{(m)H}\mathbf{r}_n\big|^2\big(\mathbf{w}_n^{(m)H}\mathbf{r}_n\big)^*\mathbf{r}_n
\end{aligned} \tag{33}$$

2) If $\gamma_n h_n(\mathbf{w}_n, \mathbf{r}_n) + \mu_n < 0$, we can similarly derive

$$\begin{aligned}
\mathbf{w}_n^{(m+1)} &= -E\{g(|y|^2)y^*\mathbf{x}\} + E\Big\{g'\big(|y|^2\big)|y|^2 + g\big(|y|^2\big)\Big\}\mathbf{w}_n^{(m)} \\
&+ E\{\mathbf{x}\mathbf{x}^T\}E\Big\{g'\big(|y|^2\big)y^{*2}\Big\}\mathbf{w}_n^{(m)*}
\end{aligned} \tag{34}$$

Combining (33) and (34), we can get

$$\begin{aligned}
\mathbf{w}_n^{(m+1)} &= -E\{g(|y|^2)y^*\mathbf{x}\} + E\Big\{g'\big(|y|^2\big)|y|^2 + g\big(|y|^2\big)\Big\}\mathbf{w}_n^{(m)} \\
&+ E\{\mathbf{x}\mathbf{x}^T\}E\Big\{g'\big(|y|^2\big)y^{*2}\Big\}\mathbf{w}_n^{(m)*} \\
&+ \text{sign}\big(\max\{\gamma_n h_n(\mathbf{w}_n, \mathbf{r}_n) + \mu_n, 0\}\big) \\
&\cdot 2\gamma_n\big|\mathbf{w}_n^{(m)H}\mathbf{r}_n\big|^2\big(\mathbf{w}_n^{(m)H}\mathbf{r}_n\big)^*\mathbf{r}_n
\end{aligned} \tag{35}$$

Updating the μ_n using the gradient descent method, we can obtain

$$\mu_n \leftarrow \max\{\gamma_n h_n(\mathbf{w}_n, \mathbf{r}_n) + \mu_n, 0\} \tag{36}$$

Competing interests
The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Author details
[1]Chongqing Key Laboratory of Nonlinear Circuits and Intelligent Information Processing, College of Electronic and Information Engineering, Southwest University, Chongqing 400715, China. [2]School of Mathematics and Statistics, Southwest University, Chongqing 400715, China.

References

1. P. Comon, Independent component analysis, a new concept? Signal Process. **36**, 287–314 (1994)
2. A. Hyvarinen, J. Karunen, E. Oja, *Independent Component Analysis* (Wiley, New York, 2001)
3. A. Tharwat, Independent component analysis: an introduction. Appl. Comput. Inform. (2018)
4. H. Saruwatari, T. Kawamura, T. Nishikawa, et al., Blind source separation based on a fast-convergence algorithm combining ICA and beamforming. IEEE Trans. Audio, Speech, and Lang. Process. **14**, 666–678 (2006)
5. K. Mohanaprasad, P. Arulmozhivarman, Wavelet-based ICA using maximum likelihood estimation and information-theoretic measure for acoustic echo cancellation during double talk situation. Circuits Systems Signal Process. **34**, 3915–3931 (2015)
6. M. Aladjem, I. Israeli-Ran, M. Bortman, Sequential independent component analysis density estimation. IEEE Trans. Neural Netw. Learn. Syst. **99**, 1–14 (2018)
7. J.F. Cardoso, A. Souloumiac, Blind beamforming for non-Gaussian signals. IEEE Proceedings F **140**, 362–370 (1993)
8. A. Yeredor, Non-orthogonal joint diagonalization in the least-squares sense with application in blind source separation. IEEE Trans. Signal Process. **50**, 1545–1553 (2002)
9. A. Hyvärinen, Fast and robust fixed-point algorithms for independent component analysis. IEEE Trans. Neural Netw. **10**, 626–634 (1999)
10. E. Bingham, A. Hyvärinen, A fast fixed-point algorithm for independent component analysis of complex valued signals. Int. J. Neural Syst. **10**, 1–8 (2000)
11. M. Novey, T. Adali, On extending the complex FastICA algorithm to noncircular sources. IEEE Trans. Signal Process. **56**, 2148–2154 (2008)
12. P. Ablin, J.F. Cardoso, A. Gramfort, Faster independent component analysis by preconditioning with Hessian approximations. IEEE Trans. Signal Process. **66**(15), 4040–4049 (2018)
13. N. Ono, S. Miyabe, in *International Conference on Latent Variable Analysis and Signal Separation.* Auxiliary-function-based independent component analysis for super-Gaussian sources (Springer, Berlin, 2010), pp. 165–172
14. S. Gepshtein, Y. Keller, Iterative spectral independent component analysis. Signal Process. **155**, 368–376 (2019)
15. N. Ono, in *Applications of Signal Processing to Audio and Acoustics (WASPAA), 2011 IEEE Workshop.* Stable and fast update rules for independent vector analysis based on auxiliary function technique (2011), pp. 189–192
16. D. Kitamura, N. Ono, H. Sawada, et al., Determined blind source separation unifying independent vector analysis and nonnegative matrix factorization. IEEE/ACM Trans. Audio Speech Lang. Proc. **24**(9), 1622–1637 (2016)
17. Y. Mitsui, N. Takamune, D. Kitamura, et al., in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).* Vectorwise coordinate descent algorithm for spatially regularized independent low-rank matrix analysis (2018), pp. 746–750
18. Y. Xia, B. Jelfs, M.M. Van Hulle, J.C. Principe, D.P. Mandic, An augmented echo state network for nonlinear adaptive filtering of complex noncircular signals. IEEE Trans. Neural Netw. **22**(1), 74–83 (2011)
19. Y. Xia, D.P. Mandic, A full mean square analysis of CLMS for second order noncircular inputs. IEEE Trans. Signal Process. **65**(21), 5578–5590 (2017)
20. W. Lu, J.C. Rajapakse, Constrained independent component analysis. Adv. Neural Inf. Process. Syst. **13**, 570–576 (2000)
21. W. Lu, J.C. Rajapakse, Approach and applications of constrained ICA. IEEE Trans. Neural Netw. **16**, 203–212 (2005)
22. J. Lee, K.L. Park, K.J. Lee, Temporally constrained ICA-based fetal ECG separation. Electron. Lett. **41**, 1158–1160 (2005)
23. J. Zhang, Z. Zhang, W. Cheng, X. Li, B. Chen, et al., Kurtosis-based constrained independent component analysis and its application on source contribution quantitative estimation. IEEE Trans. Instrum. Meas. **63**, 1842–1854 (2014)
24. P.A. Rodriguez, M. Anderson, X.L. Li, T. Adali, General non-orthogonal constrained ICA. IEEE Trans. Signal Process. **62**, 2778–2786 (2014)
25. Y. Shi, W. Zeng, N. Wang, Z. Le, A new method for independent component analysis with priori information based on multi-objective optimization. J. Neurosci. Methods **283**, 72–82 (2017)
26. X. Wang, Z. Huang, Y. Zhou, et al., Approaches and applications of semi-blind signal extraction for communication signals based on constrained independent component analysis: the complex case. Neurocomputing **101**, 204–216 (2013)
27. D. Kitamura, S. Mogami, Y. Mitsui, et al., Generalized independent low-rank matrix analysis using heavy-tailed distributions for blind source separation. EURASIP J. Adv. Sign. Proc. **28**, 1–25 (2018)
28. S. Amari, A. Cichocki, H. Yang, in *Advances in Neural Information Processing Systems*, ed. by D. S. Touretzky, M. C. Mozer, M. E. Hasselmo. A new learning algorithm for blind signal separation, vol 8 (MIT Press, Cambridge, 1996), pp. 757–763