A Novel Derivation of Nc-FastICA and Convergence Analysis of C-FastICA

Qianru Ma, Jimin Ye

Abstract—**In this paper, we utilize generalized linear (or linear-conjugation-linear) transformation to re-derive the nc-FastICA algorithm, and degenerate to obtain the c-FastICA algorithm. In addition, the necessary and sufficient conditions for the fixed point (pseudo-fixed point) of the iterative function for the c-FastICA algorithm are first proved. We also introduce in detail the relationship between the fixed point (pseudo-fixed point) and the local minima(maxima) point of contrast function in the orthogonal projection method. Finally, it is proved that the column vectors of the mixing matrix are the fixed point (pseudo-fixed point) of the iterative function. Three properties: of the c-FastICA algorithm and nc-FastICA algorithm are verified by computer simulation. (1) Both two algorithms are convergent. (2) The larger sample size, the better separation effect. (3) There is poor separation effect when source signal is selected as Gaussian distribution.**

***Index Terms*—nc-FastICA, c-FastICA, fixed point, minima, pseudo-fixed point, maxima**

I. INTRODUCTION

INDEPENDENT component analysis (ICA) [1]-[3] is a classic statistical technique that transforms mixed signals into components that are mutually as independent as possible, mainly for dealing with blind source separation (BSS) problem. ICA is widely used in many fields, such as feature extraction, medical signal processing, image signal processing, speech signal processing, fetal ECG separation, and so on. There are many existing methods for ICA. And fast independent component analysis (FastICA) [4]-[5] is extremely popular with its fast convergence speed.

Since Hyvärinen proposed the FastICA algorithm in [4], many scholars have conducted related researches. As we all know, the FastICA algorithm mainly includes two versions. (1) one-unit (or deflation) FastICA [6], which can only separate one source signal at a time; (2) symmetric FastICA, the required source signals can be separated at the same time, equivalent to parallel execution of several one-unit FastICA algorithms. Consecutively, the complex-valued FastICA algorithm [8] has also been proposed to solve the complex BSS [9]-[10]. In 2000 Bingham and Hyvärinen first proposed the complex FastICA algorithm for processing circular (proper) complex-valued signals (abbreviation c-FastICA)[8]. In 2008, Novey extended it to the noncircular complex FastICA (nc-FastICA) algorithm [11]. Due to the inherent complexity of the complex value field, the research on the complex FastICA algorithm is scarce, compared with the research of the real FastICA algorithm. Li proposed a novel complex ICA by entropy bound minimization (ICA-EBM), based on a novel (differential) entropy estimator [12]. Loesch and Yang [13] derived a closed-form expression for the Cramér-Rao bound (CRB) of the demixing matrix for instantaneous noncircular complex FastICA algorithm. In [14], Chao and Douglas proposed to use the Huber M-estimator as a nonlinearity within the complex FastICA algorithm, and demonstrated the ability of the proposed algorithm to separate mixtures of various complex-valued sources. An improved nc-FastICA algorithm is proposed for the separation of digital communication signals, and it is asymptotically efficient, i.e. its estimation error can be made much smaller by adaptively choosing the approximate optimal nonlinearity [15]. Ruan and Li [16] discussed two complex FastICA algorithms for noisy data, where contrast functions are based on kurtosis and negentropy respectively, also gave the stability conditions of contrast functions. In [17], a simple check of undesirable points was proposed based on the kurtosis of estimated sources to get rid of the undesirable fixed points. The authors of [18] established the theory for complex-valued FastICA giving Cramér -Rao lower bound and identification conditions, and presented a new algorithm that taken three properties into account, including non-Gaussianity, nonwhiteness, and noncircularity. In [19], Zhao W *et al* developed a new reference-based contrast function by introducing reference signals into the negentropy, upon which an efficient optimal FastICA algorithm is derived for noncircular sources. Related study on complex valued FastICA is available in [20]-[21].

The first major purpose of this article is to re-derive the nc-FastICA algorithm in a simpler method. In [8], the circular complex-valued FastICA algorithm (c-FastICA) is discussed, and locally stable conditions are given which is analogous to real-valued FastICA algorithm [4]. In 2008, Novey and Adali [11] derived the nc-FastICA algorithm and local stability conditions. In this paper, we utilize generalized linear (or linear-conjugation-linear) transformation [22] and re-derive the nc-FastICA algorithm .

The convergence of the real-valued FastICA algorithm has been discussed in [23]-[24]. For the real-valued FastICA algorithm author of [24] proved the minimums of contrast

Manuscript received March 15th, 2019; revised May 20th, 2019.

This work was supported in part by the National Natural Science Foundation of China under Grant 61573014, Central University Basic Research Fund under Grant JB180702.

Qianru Ma is q master student at Xidian University, Xi'an China, mainly study independent component analysis, blind signal processing(e-mail: mqr213237@163.com).

Jimin Ye is a professor in Xidian University, Xi'an China. His scientific interests are risk theory, statistical signal processing, and blind source separation using independent component analysis. (corresponding author phone:13991301192 e-mail: jmye@mail.xidian.edu.cn).

function are fixed points of FastICA iterative function. Furthermore, the convergence order of the algorithm was proved. However, it is still a missing for the complex-valued FastICA algorithm. In this paper, we fills this gap. In the part IV, we introduce in detail the relationship between the fixed point (pseudo-fixed point) of iterative function and the local minima point of contrast function in the orthogonal projection method.

II. Complex ICA

A. Preliminaries

In this paper, we define complex random vector $\boldsymbol{x} = \boldsymbol{x}_R + j\boldsymbol{x}_I \in C^n$,where $\boldsymbol{x}_R$ and $\boldsymbol{x}_I$ denote the real and imaginary parts of $\boldsymbol{x}$ respectively, and $j^2 = -1$. We introduce augmented vector $\underline{\boldsymbol{x}}$ of $\boldsymbol{x}$, and the specific calculation formula as follows,

$$\underline{\boldsymbol{x}} \triangleq \begin{pmatrix} \boldsymbol{x} \\ \boldsymbol{x}^* \end{pmatrix} = \boldsymbol{T} \begin{pmatrix} \boldsymbol{x}_R \\ \boldsymbol{x}_I \end{pmatrix} = \begin{pmatrix} \boldsymbol{I}_n & j\boldsymbol{I}_n \\ \boldsymbol{I}_n & -j\boldsymbol{I}_n \end{pmatrix} \begin{pmatrix} \boldsymbol{x}_R \\ \boldsymbol{x}_I \end{pmatrix},$$

where $\boldsymbol{x}^*$ denotes the conjugate of $\boldsymbol{x}$. $\boldsymbol{T} \in C^{2n\times 2n}$ is a transformation matrix from real value to complex value, and $\boldsymbol{T}^H\boldsymbol{T} = \boldsymbol{T}\boldsymbol{T}^H = 2\boldsymbol{I}_{2n}$, where $\boldsymbol{T}^H$ is the conjugate transpose of $\boldsymbol{T}$. Obviously, $\begin{pmatrix} \boldsymbol{x}_R & \boldsymbol{x}_I \end{pmatrix}^T \in R^{2n}$, where T denotes the transpose. Write the complex space to which $\underline{\boldsymbol{x}}$ belongs as C_*^{2n}, whose first n elements are the conjugate of the last n elements. The symbols associated with the augmented vectors mentioned in this article are underlined. The expectation of $\underline{\boldsymbol{x}}$:

$$\underline{\boldsymbol{\mu}}_x = E\{\underline{\boldsymbol{x}}\} = \begin{pmatrix} \boldsymbol{\mu}_x & \boldsymbol{\mu}_{x^*} \end{pmatrix}^T = \begin{pmatrix} \boldsymbol{\mu}_{x_R} + j\boldsymbol{\mu}_{x_I} & \boldsymbol{\mu}_{x_R} - j\boldsymbol{\mu}_{x_I} \end{pmatrix}^T.$$

And augmented covariance matrix of $\underline{\boldsymbol{x}}$ is computed as follows,

$$\underline{\boldsymbol{R}}_{xx} = E\{(\underline{\boldsymbol{x}} - \underline{\boldsymbol{\mu}}_x)(\underline{\boldsymbol{x}} - \underline{\boldsymbol{\mu}}_x)^H\} = \begin{pmatrix} \boldsymbol{R}_{xx} & \tilde{\boldsymbol{R}}_{xx} \\ \tilde{\boldsymbol{R}}_{xx}^* & \boldsymbol{R}_{xx}^* \end{pmatrix}, \quad (1)$$

where $\boldsymbol{R}_{xx} = E\{(\boldsymbol{x} - \boldsymbol{\mu}_x)(\boldsymbol{x} - \boldsymbol{\mu}_x)^H\} = E\{\boldsymbol{x}\boldsymbol{x}^H\}$ is a Hermit matrix, and $\tilde{\boldsymbol{R}}_{xx} = E\{(\boldsymbol{x} - \boldsymbol{\mu}_x)(\boldsymbol{x} - \boldsymbol{\mu}_x)^T\} = E\{\boldsymbol{x}\boldsymbol{x}^T\}$ called pseudo-covariance matrix is a symmetric matrix. If $\tilde{\boldsymbol{R}}_{xx} = \boldsymbol{0}$, random vector $\boldsymbol{x}$ is circular or proper. Otherwise, it is noncircular.

B. Complex-valued ICA

The mathematical model of complex-valued ICA is defined by

$$\boldsymbol{x} = \boldsymbol{A}\boldsymbol{s}. \quad (2)$$

where $\boldsymbol{s} = [s_1, s_2, \cdots, s_n]^T$ denotes unknown source signal to be recovered, whose components $s_i (i = 1, \cdots, n)$ are mutually independent. $\boldsymbol{x} = [x_1, x_2, \cdots, x_n]^T$ denotes mixed signal that can be observed. $\boldsymbol{A}$ is an unknown ($n \times n$) mixing matrix. For simplicity, following hypotheses are made:

Assumption 1: Source signal $\boldsymbol{s}$ satisfies $E\{\boldsymbol{s}\boldsymbol{s}^H\} = \boldsymbol{I}$.

Assumption 2: $\boldsymbol{x}$ has been performed whitening and centering process, i.e. $E\{\boldsymbol{x}\boldsymbol{x}^H\} = \boldsymbol{I}$ and $\boldsymbol{\mu}_x = \boldsymbol{0}$.

Assumption 3: $\boldsymbol{A}$ is square and non-singular with full rank. Without loss of generality we also assume $\boldsymbol{A}$ is a unitary matrix as described in [11].

For circular source signal, $E\{\boldsymbol{s}\boldsymbol{s}^T\} = \boldsymbol{0}$, and $E\{\boldsymbol{s}\boldsymbol{s}^T\} \neq \boldsymbol{0}$ for noncircular source signal. Introduction for the whitening of complex-valued signals can be found in [25]. We aim to look for the optimal separated matrix $\boldsymbol{W}$ to make $y = \boldsymbol{w}^H\boldsymbol{x}$ as an estimate of the source signal, where $\boldsymbol{w}^H$ is row vector of $\boldsymbol{W}$ and is referred as separated vector. Of course, $\boldsymbol{W}$ is a unitary matrix [11].

III. Noncircular Complex FastICA

Similar to real-value FastICA algorithm, complex-valued FastICA algorithm is based on solving the following optimization,

$$\hat{\boldsymbol{w}} = \arg\max_{\boldsymbol{w}\in S^{n-1}} E\{G(|\boldsymbol{w}^H\boldsymbol{x}|^2)\}. \quad (3)$$

$G(\cdot): R^+ \cup \{0\} \to R$ is a twice continuously differentiable and non-quadratic function, which is referred to as the nonlinearity, and $S^{n-1} = \{\boldsymbol{w} \in C^n \,\|\, \|\boldsymbol{w}\|^2 = 1\}$. In addition, $J(\boldsymbol{w}) = E\{G(|\boldsymbol{w}^H\boldsymbol{x}|^2)\}$ is called contrast (cost) function. There have been three different classic nonlinearities [8] and derivatives $g(\cdot)$:

$$G_1(y) = \sqrt{a_1 + y}\,,\; g_1(y) = \frac{1}{2\sqrt{a_1 + y}}\,,$$

$$G_2(y) = \log(a_2 + y)\,,\; g_2(y) = \frac{1}{a_2 + y}\,,$$

$$G_3(y) = \frac{1}{2}y^2\,,\; g_3(y) = y\,,$$

where a_1 and a_2 are arbitrary constants whose values usually are chosen as 0.1. In [11], the nc-FastICA algorithm has been derived. In this paper, we re-derive this algorithm by using a simple method, generalized linear transformation [22] (or linear-conjugation-linear).

The complex differential theory used in this paper is Wirtinger differential [26]-[27], whose main idea is to treat f as a function of two independent complex variables $\boldsymbol{x}$ and $\boldsymbol{x}^*$. Thus, solving a generalized complex differential with respect to $\boldsymbol{x}$, we treats $\boldsymbol{x}^*$ as a constant theoretically. The generalized complex gradient is defined as:

$$\nabla_x f = (\frac{\partial f}{\partial \boldsymbol{x}}, \frac{\partial f}{\partial \boldsymbol{x}^*})^T, \quad (4)$$

where generalized complex differential

$$\frac{\partial}{\partial \boldsymbol{x}} \triangleq \frac{1}{2}(\frac{\partial}{\partial \boldsymbol{x}_R} - j\frac{\partial}{\partial \boldsymbol{x}_I}), \quad (5)$$

and conjugate generalized complex differential

$$\frac{\partial}{\partial \boldsymbol{x}^*} \triangleq \frac{1}{2}(\frac{\partial}{\partial \boldsymbol{x}_R} + j\frac{\partial}{\partial \boldsymbol{x}_I}). \quad (6)$$

$\boldsymbol{x}$ of (4)-(6) can be either a vector or a scalar, so can f . Specific extensions can be found in [26]-[27]. If f is a real value function, $f(\boldsymbol{x}) = f^*(\boldsymbol{x})$. Thus, the gradient of the scalar real function $f(\boldsymbol{x})$ is the augmented vector

$$\nabla_x f = (\frac{\partial f}{\partial \boldsymbol{x}}, (\frac{\partial f}{\partial \boldsymbol{x}})^*)^T. \tag{7}$$

The generalized complex Hessian matrix is defined as:

$$\underline{\boldsymbol{H}}_{xx} = \frac{\partial f}{\partial \underline{\boldsymbol{x}}}(\frac{\partial f}{\partial \underline{\boldsymbol{x}}})^H = \begin{pmatrix} \boldsymbol{H}_{xx} & \tilde{\boldsymbol{H}}_{xx} \\ \tilde{\boldsymbol{H}}_{xx}^* & \boldsymbol{H}_{xx}^* \end{pmatrix}, \tag{8}$$

where complex Hessian matrix

$$\boldsymbol{H}_{xx} = \frac{\partial}{\partial \boldsymbol{x}}(\frac{\partial f}{\partial \boldsymbol{x}})^H, \tag{9}$$

and complex complementary Hessian matrix

$$\tilde{\boldsymbol{H}}_{xx} = \frac{\partial}{\partial \boldsymbol{x}^*}(\frac{\partial f}{\partial \boldsymbol{x}})^H. \tag{10}$$

Next, we re-derive the nc-FastICA algorithm using generalized linear transformation (or linear-conjugate-linear) and the above differential theory. We solve constraint optimization problem (3) based on Lagrangian multiplier. First, construct Lagrangian function

$$L(\boldsymbol{w}, \lambda) = J(\boldsymbol{w}) + \beta(\boldsymbol{w}^H \boldsymbol{w} - 1), \tag{11}$$

where β is Lagrangian multiplier. Then, solve (11) based on complex Newton's update [27],

$$\Delta \underline{\boldsymbol{w}} = -(\underline{\boldsymbol{H}}_{ww})^{-1} \nabla_{w^*} J, \tag{12}$$

where $\Delta \underline{\boldsymbol{w}} = \underline{\boldsymbol{w}}_{n+1} - \underline{\boldsymbol{w}}_n$. Substitute (11) into (12),

$$\Delta \underline{\boldsymbol{w}} = -(\underline{\boldsymbol{H}}_{ww} J + \lambda \boldsymbol{I}_{2n})^{-1} (\nabla_w J + \lambda \underline{\boldsymbol{w}}).$$

where $\lambda = 2\beta$. Further, we have

$$(\underline{\boldsymbol{H}}_{ww} J + \lambda \boldsymbol{I}_{2n}) \underline{\boldsymbol{w}} \leftarrow -\nabla_{w^*} J + \underline{\boldsymbol{H}}_{ww} J \underline{\boldsymbol{w}}. \tag{13}$$

According to (5), (6), and (7), we obtain the generalized complex gradient as

$$\nabla_w J = \begin{pmatrix} \frac{\partial J}{\partial \boldsymbol{w}_1} & \cdots & \frac{\partial J}{\partial \boldsymbol{w}_n} & \frac{\partial J}{\partial \boldsymbol{w}_1^*} & \cdots & \frac{\partial J}{\partial \boldsymbol{w}_n^*} \end{pmatrix}^T = \begin{pmatrix} E\{g(yy^*)y\boldsymbol{x}_1^*\} \\ \vdots \\ E\{g(yy^*)y\boldsymbol{x}_n^*\} \\ E\{g(yy^*)y^*\boldsymbol{x}_1\} \\ \vdots \\ E\{g(yy^*)y^*\boldsymbol{x}_n\} \end{pmatrix}. \tag{14}$$

From (8), if solving the generalized complex Hessian matrix $\underline{\boldsymbol{H}}_{ww} J$, we need to find $\boldsymbol{H}_{ww} J$ and $\tilde{\boldsymbol{H}}_{ww} J$. Elements in the matrix $\boldsymbol{H}_{ww} J$ is

$$\frac{\partial^2 J}{\partial \boldsymbol{w}_k \partial \boldsymbol{w}_i^*} = E\{\boldsymbol{x}_i \boldsymbol{x}_k^* [g'(yy^*)yy^* + g(yy^*)]\}, \tag{15}$$

where $g'(\cdot)$ is derivative of function $g(\cdot)$. And elements in the matrix $\tilde{\boldsymbol{H}}_{ww} J$ is

$$\frac{\partial^2 J}{\partial \boldsymbol{w}_k^* \partial \boldsymbol{w}_i^*} = E\{\boldsymbol{x}_i \boldsymbol{x}_k g'(yy^*) y^{*2}\}. \tag{16}$$

Thus, the generalized complex Hessian matrix

$$\underline{\boldsymbol{H}}_{ww} J = E\begin{pmatrix} \boldsymbol{x}\boldsymbol{x}^H c & \boldsymbol{x}\boldsymbol{x}^T d \\ \boldsymbol{x}^*\boldsymbol{x}^H d & \boldsymbol{x}^*\boldsymbol{x}^T c \end{pmatrix}, \tag{17}$$

where $c = g'(|y|^2)|y|^2 + g(|y|^2)$ and $d = g'(|y|^2)y^{*2}$. According to the part 2.2.1 of [22], we know $\underline{\boldsymbol{H}}_{ww} J \in W^{n\times n}$, where $W^{n\times n}$ denotes a set composed of $2n \times 2n$ augmented matrices $\underline{\boldsymbol{H}}$ that satisfy (18), i.e. a specific modular form: the lower right matrix block is the conjugate of the upper left matrix block, and the lower left matrix block is the conjugate of the upper right matrix block.

$$\underline{\boldsymbol{H}} = \begin{pmatrix} \boldsymbol{H}_1 & \boldsymbol{H}_2 \\ \boldsymbol{H}_2^* & \boldsymbol{H}_1^* \end{pmatrix}. \tag{18}$$

Then from (2.10) of [22], we can obtain a augmented description (19) of generalized linear transformation (or linear-conjugate-linear),

$$\underline{\boldsymbol{H}}_{ww} J\boldsymbol{w} = E\{\boldsymbol{x}\boldsymbol{x}^H c\}\boldsymbol{w} + E\{\boldsymbol{x}\boldsymbol{x}^T d\}\boldsymbol{w}^*. \tag{19}$$

Taking the first n rows of (13), then combining (14) and (19), we have

$$\boldsymbol{w} \leftarrow -\frac{\partial J}{\partial \boldsymbol{w}^*} + E\{\boldsymbol{x}\boldsymbol{x}^H c\}\boldsymbol{w} + E\{\boldsymbol{x}\boldsymbol{x}^T d\}\boldsymbol{w}^*. \tag{20}$$

Equation (20) removes the coefficient of the left $\boldsymbol{w}$, which is reasonable from [11]. And $E\{\boldsymbol{x}\boldsymbol{x}^H c\} \approx E\{\boldsymbol{x}\boldsymbol{x}^H\}c$, $E\{\boldsymbol{x}\boldsymbol{x}^T d\} \approx E\{\boldsymbol{x}\boldsymbol{x}^T\}d$ [11]. Thus, nc-FastICA update for one unit is

$$\begin{cases} \boldsymbol{w} \leftarrow -E\{g(|y|^2)y^*\boldsymbol{x}\} + E\{g'(|y|^2)|y|^2 + g(|y|^2)\}\boldsymbol{w} \\ \quad + E\{\boldsymbol{x}\boldsymbol{x}^T\}E\{g'(|y|^2)y^{*2}\}\boldsymbol{w}^* \\ \boldsymbol{w} \leftarrow \frac{\boldsymbol{w}}{\|\boldsymbol{w}\|} \end{cases}. \tag{21}$$

$g(\cdot)$ and $g'(\cdot)$ in the iteration (21) represent the first derivative and the second derivative of the nonlinear $G(\cdot)$, respectively. Of course, c-FastICA update for one unit is

$$\begin{cases} \boldsymbol{w} \leftarrow -E\{g(|y|^2)y^*\boldsymbol{x}\} + E\{g'(|y|^2)|y|^2 + g(|y|^2)\}\boldsymbol{w} \\ \boldsymbol{w} \leftarrow \frac{\boldsymbol{w}}{\|\boldsymbol{w}\|} \end{cases}. \tag{22}$$

IV. CONVERGENCE FOR C-FASTICA

In recent years, many scholars have studied the convergence of the real-valued FastICA algorithm. Specific, Oja, Yuan and Shen, Kleinsteuber, Huper discussed the convergence of the real-valued FastICA algorithm in [23]-[24]. Especially, author proved the minimum of the real-valued FastICA algorithm is a fixed point of the FastICA iterative function In [24]. However, this is still a missing for the complex-valued FastICA algorithm. Thus, we discussed convergence for the c-FastICA algorithm.

Firstly, based on model (22) and the local stability conditions of the c-FastICA algorithm, we define four important functions for relevant explanations and proofs below:

A. $R: S^{n-1} \to C^n$

$$\begin{aligned} R(\boldsymbol{w}) := & -E\{g(|\boldsymbol{w}^H\boldsymbol{x}|^2)(\boldsymbol{w}^H\boldsymbol{x})^*\boldsymbol{x}\} \\ & + E\{g'(|\boldsymbol{w}^H\boldsymbol{x}|^2)|\boldsymbol{w}^H\boldsymbol{x}|^2 + g(|\boldsymbol{w}^H\boldsymbol{x}|^2)\}\boldsymbol{w} \end{aligned} \tag{23}$$

B. $T: S^{n-1} \to C^n$

$$T(\boldsymbol{w}) := \frac{R(\boldsymbol{w})}{\|R(\boldsymbol{w})\|} \tag{24}$$

C. $U: S^{n-1} \to C$

$$\begin{aligned} U(w) := & E\{g(|\boldsymbol{w}^H\boldsymbol{x}|^2) + |\boldsymbol{w}^H\boldsymbol{x}|^2 g'(|\boldsymbol{w}^H\boldsymbol{x}|^2) \\ & - |\boldsymbol{w}^H\boldsymbol{x}|^2 g(|\boldsymbol{w}^H\boldsymbol{x}|^2)\} \end{aligned} \tag{25}$$

D. $\Psi : S^{n-1} \to C$

$$\Psi(\boldsymbol{v}) := E\{(\boldsymbol{v}^H\boldsymbol{x})^* g(|\boldsymbol{v}^H\boldsymbol{x}|^2)\boldsymbol{P}_v^{\perp}\boldsymbol{x}\} \tag{26}$$

And C^n denotes all n dimensions complex vectors. Actually, $R(\boldsymbol{w})$ and $T(\boldsymbol{w})$ is consistent with (22). $U(\boldsymbol{w})$ is convergence condition, this can be found in [8]. If $U(\boldsymbol{w}) < 0$, the column of the mixing matrix $\boldsymbol{A}$ is local maxima of contrast function under the constraint $E\{|\boldsymbol{w}^H\boldsymbol{x}|^2\} = \|\boldsymbol{w}\|^2 = 1$. Otherwise, they are local minima. Equation (26) will be used to prove theorem 1 and will be used for judging criteria of fixed point of the c-FastICA algorithm.

In following, we use orthogonal projection method to analyze the convergence of the c-FastICA algorithm. We assume $\boldsymbol{P}_w$ denote matrix of orthogonal projection from C^n to subspace $\text{span}(\boldsymbol{w})$, of course, $\boldsymbol{w} \in S^{n-1}$. And $\boldsymbol{P}_w^{\perp}$ is matrix of orthogonal projection from C^n to subspace $\text{span}(\boldsymbol{w})^{\perp}$. Then, we can know

$$\boldsymbol{P}_w = \boldsymbol{w}\boldsymbol{w}^H \tag{27}$$

and

$$\boldsymbol{P}_w^{\perp} = \boldsymbol{I} - \boldsymbol{w}\boldsymbol{w}^H \tag{28}$$

For any $\boldsymbol{x} \in C^n$, the following orthogonal decomposition is effective.

$$\boldsymbol{x} = \boldsymbol{P}_w\boldsymbol{x} + \boldsymbol{P}_w^{\perp}\boldsymbol{x} \tag{29}$$

Next, we discuss some relative convergence problems on c-FastICA algorithm. Theorem 1 is based on definition 1 and the orthogonal decomposition of vector $\boldsymbol{v}$. And definition 1 is displayed as follows.

Definition 1 If $\boldsymbol{x}$ satisfy $f(\boldsymbol{x}) = -\boldsymbol{x}$, $\boldsymbol{x}$ is referred as pseudo-fixed point.

Theorem 1 (Fixed (Pseudo-fixed) point condition) The necessary and sufficient condition for the vector $\boldsymbol{v}$ to be a fixed point of the iterative function (24) is $\Psi(\boldsymbol{v}) = 0$ and $U(\boldsymbol{v}) > 0$. And the necessary and sufficient condition for the vector $\boldsymbol{v}$ to be a pseudo-fixed point of the iterative function (24) is $\Psi(\boldsymbol{v}) = 0$ and $U(\boldsymbol{v}) < 0$.

Proof: According to the definition of fixed point, we know the vector $\boldsymbol{v}$ is a fixed point of the iterative function $T(\boldsymbol{w})$ if and only if $T(\boldsymbol{v}) = \boldsymbol{v}$.

$$\begin{aligned} T(\boldsymbol{v}) &= \frac{1}{\|R(\boldsymbol{v})\|}\Big[-E\{g(|\boldsymbol{v}^H\boldsymbol{x}|^2)(\boldsymbol{v}^H\boldsymbol{x})^*\boldsymbol{x}\} \\ &\quad + E\{g'(|\boldsymbol{v}^H\boldsymbol{x}|^2)|\boldsymbol{v}^H\boldsymbol{x}|^2 + g(|\boldsymbol{v}^H\boldsymbol{x}|^2)\}\boldsymbol{v}\Big] \\ &= \frac{1}{\|R(\boldsymbol{v})\|}\Big[E\{g'(|\boldsymbol{v}^H\boldsymbol{x}|^2)|\boldsymbol{v}^H\boldsymbol{x}|^2 + g(|\boldsymbol{v}^H\boldsymbol{x}|^2)\}\boldsymbol{v} \\ &\quad - E\{g(|\boldsymbol{v}^H\boldsymbol{x}|^2)(\boldsymbol{v}^H\boldsymbol{x})^*(\boldsymbol{P}_v\boldsymbol{x} + \boldsymbol{P}_v^{\perp}\boldsymbol{x})\}\Big] \\ &= \frac{1}{\|R(\boldsymbol{v})\|}\Big[E\{g'(|\boldsymbol{v}^H\boldsymbol{x}|^2)|\boldsymbol{v}^H\boldsymbol{x}|^2 + g(|\boldsymbol{v}^H\boldsymbol{x}|^2)\}\boldsymbol{v} \\ &\quad - E\{|\boldsymbol{v}^H\boldsymbol{x}|^2 g(|\boldsymbol{v}^H\boldsymbol{x}|^2)\boldsymbol{v}\} - E\{g(|\boldsymbol{v}^H\boldsymbol{x}|^2)(\boldsymbol{v}^H\boldsymbol{x})^*\boldsymbol{P}_v^{\perp}\boldsymbol{x}\}\Big] \\ &= \frac{1}{\|R(\boldsymbol{v})\|}\big[U(\boldsymbol{v})\boldsymbol{v} - \Psi(\boldsymbol{v})\big]. \end{aligned}$$

Obviously $\Psi(\boldsymbol{v}) = E\{(\boldsymbol{v}^H\boldsymbol{x})^* g(|\boldsymbol{v}^H\boldsymbol{x}|^2)\boldsymbol{P}_v^{\perp}\boldsymbol{x}\} \in \text{span}(\boldsymbol{v})^{\perp}$. And it is orthogonal to $\boldsymbol{v}$. Thus, $T(\boldsymbol{v})$ is parallel with $\boldsymbol{v}$ if and only if $\Psi(\boldsymbol{v}) = 0$.

Next we have

$$\begin{aligned} \|R(\boldsymbol{v})\| = \|U(\boldsymbol{v})\boldsymbol{v} - \Psi(\boldsymbol{v})\| &= [\|U(\boldsymbol{v})\boldsymbol{v}\|^2 + \|\Psi(\boldsymbol{v})\|^2 \\ &\quad - \langle U(\boldsymbol{v})\boldsymbol{v}, \Psi(\boldsymbol{v})\rangle - \langle \Psi(\boldsymbol{v}), U(\boldsymbol{v})\boldsymbol{v}\rangle]^{\frac{1}{2}} \\ &= \|U(\boldsymbol{v})\boldsymbol{v}\| = |U(\boldsymbol{v})\boldsymbol{v}|, \end{aligned}$$

where $\langle U(\boldsymbol{v})\boldsymbol{v}, \Psi(\boldsymbol{v})\rangle$ denotes inner product of $U(\boldsymbol{v})\boldsymbol{v}$ and $\Psi(\boldsymbol{v})$. Because $U(\boldsymbol{v})\boldsymbol{v} \in \text{span}(\boldsymbol{v})$ and $\Psi(\boldsymbol{v}) \in \text{span}(\boldsymbol{v})^{\perp}$, then $\langle U(\boldsymbol{v})\boldsymbol{v}, \Psi(\boldsymbol{v})\rangle = 0$. When $U(\boldsymbol{v}) > 0$, we have $T(\boldsymbol{v}) = \boldsymbol{v}$, and $\boldsymbol{v}$ a is fixed point. When $U(\boldsymbol{v}) < 0$, we have $T(\boldsymbol{v}) = -\boldsymbol{v}$. From definition 1, we know $\boldsymbol{v}$ is a pseudo-fixed point at this case.

According to Theorem 1 and the following lemma 1 [8], we can directly get the theorem 2. The specific content of Lemma 1 is as follows

Lemma 1 (convergence condition) Assume that observed signals $\boldsymbol{x}$ is prewhitened using matrix $E\{\boldsymbol{x}\boldsymbol{x}^H\} = \boldsymbol{I}$. Also, $G(\cdot)$ is a sufficient smooth even function. Then the local maxima (resp. minima) of $J(\boldsymbol{w})$ under the constraint $\|\boldsymbol{w}\|^2 = 1$ include those columns of the mixing matrix $\boldsymbol{A}$ such that the corresponding independent components s_i satisfy

$$E\{g(|s_i|^2) + |s_i|^2 g'(|s_i|^2) - |s_i|^2 g(|s_i|^2)\} < 0 (> 0, \text{ resp.}). \tag{30}$$

Equation (30) is consistent with $U(\boldsymbol{a}_i)$, where $\boldsymbol{a}_i$ is the *ith* column of mixing matrix $\boldsymbol{A}$ and $s_i = \boldsymbol{a}_i^H\boldsymbol{x}$. There is specific proof of lemma 1 in [8]. According to Theorem 1 and the following lemma 1 [8], we have following theorem 2.

Theorem 2 If $\boldsymbol{v}$ is a fixed point of the iterative function (24), then $\boldsymbol{v}$ is the local minima of the contrast function $J(\boldsymbol{w})$. Conversely, if $\boldsymbol{v}$ is a pseudo-fixed point of the iterative function (24), then $\boldsymbol{v}$ is the local maxima of the contrast function $J(\boldsymbol{w})$.

Obviously, theorem 2 is correct from theorem 1 and lemma 1. Further, we derive theorem 3 for the case of the c-FastICA algorithm. To prove theorem 3, first give the following lemma 2.

Lemma 2 Assume $\boldsymbol{a}$ is the *ith* column of mixing matrix $\boldsymbol{A}$. Then, $\boldsymbol{P}_a\boldsymbol{x}$ and $\boldsymbol{P}_a^{\perp}\boldsymbol{x}$ are independent.

Proof: Mixing matrix $\boldsymbol{A}$ is unitary matrix, i.e. $\boldsymbol{A}\boldsymbol{A}^H = \boldsymbol{A}^H\boldsymbol{A} = \boldsymbol{I}$. Then, we have $\boldsymbol{a}^H\boldsymbol{A} = \boldsymbol{e}_i^T$. Further, $\boldsymbol{a}^H\boldsymbol{x} = \boldsymbol{a}^H\boldsymbol{A}\boldsymbol{s} = \boldsymbol{e}_i^T\boldsymbol{s} = s_i$.

Because $\boldsymbol{x} = \boldsymbol{A}\boldsymbol{s} = \sum_{j=1}^{n}\boldsymbol{a}_j s_j$ and (28), (29), then

$$\boldsymbol{P}_a^{\perp}\boldsymbol{x} = (\boldsymbol{I} - \boldsymbol{a}\boldsymbol{a}^H)\boldsymbol{x} = \boldsymbol{x} - \boldsymbol{a}\boldsymbol{a}^H\boldsymbol{x} = \sum_{j=1}^{n}\boldsymbol{a}_j s_j - \boldsymbol{a}_i s_i = \sum_{j \neq i}\boldsymbol{a}_j s_j,$$

$$\boldsymbol{P}_a\boldsymbol{x} = \boldsymbol{a}\boldsymbol{a}^H\boldsymbol{x} = \boldsymbol{a}s_i.$$

According to s_i and s_j $(i \neq j)$ are mutually independent, so $\boldsymbol{P}_a\boldsymbol{x}$ and $\boldsymbol{P}_a^{\perp}\boldsymbol{x}$ are also independent.

Lemma 2 indicates that the orthogonal decomposition $\boldsymbol{P}_a\boldsymbol{x}$ and $\boldsymbol{P}_a^{\perp}\boldsymbol{x}$ of vector $\boldsymbol{x}$ are independent each other. Applying the conclusion of theorem 2 to the c-FastICA algorithm, the following theorem 3 is obtained.

Theorem 3 Assume $\boldsymbol{a}$ is the *ith* column of mixing matrix $\boldsymbol{A}$, and $U(\boldsymbol{a})>0$. Then, $\boldsymbol{a}$ is a fixed point of the iterative function (24) and the local minima of contrast function. Otherwise, $\boldsymbol{a}$ is a pseudo-fixed point of the iterative function (24) and the local maxima of contrast function.

Proof: From lemma 1, we can know column vector $\boldsymbol{a}_j$ ($j=1,\cdots,n$) are the local minima or maxima of the contrast function. Then we just need to prove $\boldsymbol{a}$ is a fixed point (or pseudo-fixed point) of (24). We will prove it in two methods in this paper.

Method 1 According to theorem 1, we need to prove $\Psi(\boldsymbol{a})=0$. Next, we will use lemma 2 to continue our proof. Then, we can obtain

$$\begin{aligned}\Psi(\boldsymbol{a}) &= E\{(\boldsymbol{a}^H\boldsymbol{x})^* g(|\boldsymbol{a}^H\boldsymbol{x}|^2)\boldsymbol{P}_a^{\perp}\boldsymbol{x}\}\\ &= E\{(\boldsymbol{a}^H\boldsymbol{x})^* g(|\boldsymbol{a}^H\boldsymbol{x}|^2)(\boldsymbol{I}-\boldsymbol{a}\boldsymbol{a}^H)\boldsymbol{x}\}\\ &= E\{(\boldsymbol{a}^H\boldsymbol{x})^* g(|\boldsymbol{a}^H\boldsymbol{x}|^2)\}E\{(\boldsymbol{I}-\boldsymbol{a}\boldsymbol{a}^H)\boldsymbol{x}\}\\ &=0.\end{aligned}$$

The above formula is established based on $E\{\boldsymbol{x}\}=0$.

Method 2: We have $\boldsymbol{A}\boldsymbol{e}_i=\boldsymbol{a}_i=\boldsymbol{a}$. Then,

$$\begin{aligned}R(\boldsymbol{a}) &= E\{g(|\boldsymbol{a}^H\boldsymbol{x}|^2)+|\boldsymbol{a}^H\boldsymbol{x}|^2 g'(|\boldsymbol{a}^H\boldsymbol{x}|^2)\}\boldsymbol{a}-E\{\boldsymbol{x}(\boldsymbol{a}^H\boldsymbol{x})^* g(|\boldsymbol{a}^H\boldsymbol{x}|^2)\}\\ &= E\{g(|s_i|^2)+|s_i|^2 g'(|s_i|^2)\}\boldsymbol{a}-E\{\boldsymbol{x}s_i^* g(|s_i|^2)\}\\ &= \boldsymbol{A}E\{g(|s_i|^2)+|s_i|^2 g'(|s_i|^2)\}\boldsymbol{e}_i-\boldsymbol{A}E\{\boldsymbol{s}s_i^* g(|s_i|^2)\}\\ &= \boldsymbol{A}E\{g(|s_i|^2)+|s_i|^2 g'(|s_i|^2)\}\boldsymbol{e}_i-\boldsymbol{A}E\{|s_i|^2 g(|s_i|^2)\}\boldsymbol{e}_i\\ &= U(\boldsymbol{a})\boldsymbol{a}.\end{aligned}$$

And, $T(a)=\dfrac{R(\boldsymbol{a})}{\|R(\boldsymbol{a})\|}=\dfrac{U(\boldsymbol{a})\boldsymbol{a}}{\|U(\boldsymbol{a})\boldsymbol{a}\|}=\boldsymbol{a}$, which is because $\|\boldsymbol{a}\|=1$.

In summary, Theorem 1, 2, and 3 show the relationship between the fixed point (pseudo-fixed point) of iterative function and the local minima point of contrast function in detail for the first time.

V. Computer Simulation

In this part three experiments are implement to verify the relevant performances of the c-FastICA algorithm and the nc-FastICA algorithm.

The first experiment is conducted to test the convergence performance of the c-FastICA algorithm and the nc-FastICA algorithm. For the c-FastICA algorithm, four complex signals $s_j=r_j(\cos f_j+i\sin f_j)$, $j=1,2,3,4$, are processed. The radius r_j of every source signals s_j is random number from four different distribution: Binomial distribution, Gamma distribution, Poisson distribution, Hypergeometric distribution, respectively. And all the parameters in these distributions are randomly generated. The phase angle f_j is uniformly distributed on $[-\pi,\pi]$. And every source signal is standardized to unite variance. The number of samples selected for this experiment is 60000. In addition the operating software selected in this article is MATLAB2017b.

TABLE I
C-FASTICA ALGORITHM

S1. Whiten $\boldsymbol{x}$ to make $E\{\boldsymbol{x}\boldsymbol{x}^H\}=\boldsymbol{I}$ and Initialize $\boldsymbol{W}$
S2. For $i=0,1,\ldots,n$
For $j=0,1,\ldots,n$

$$\begin{aligned}\boldsymbol{W}(:,j) &= -E\{g(|\boldsymbol{W}(:,j)|^2(\boldsymbol{W}(:,j))^H\boldsymbol{x})^*)\}\\ &\quad +E\{g'(|\boldsymbol{W}(:,j)^H\boldsymbol{x}|^2)|\boldsymbol{W}(:,j)^H\boldsymbol{x}|^2\\ &\quad +g(|\boldsymbol{W}(:,j)^H\boldsymbol{x}|^2)\}\boldsymbol{W}(:,j)\end{aligned}$$

end
$\boldsymbol{W}=(\boldsymbol{W}\boldsymbol{W}^H)^{1/2}\boldsymbol{W}$
end
S4. $\boldsymbol{s}=\boldsymbol{W}^H\boldsymbol{x}$

The mixing matrix $\boldsymbol{A}$ is a random complex matrix, and both real part and imaginary part are randomly generated. $\boldsymbol{Q}$ is whitening matrix. $\boldsymbol{W}$ is separated matrix so that $\boldsymbol{s}=\boldsymbol{W}^H\boldsymbol{x}$. The c-FastICA algorithm can be summarized as Table I.

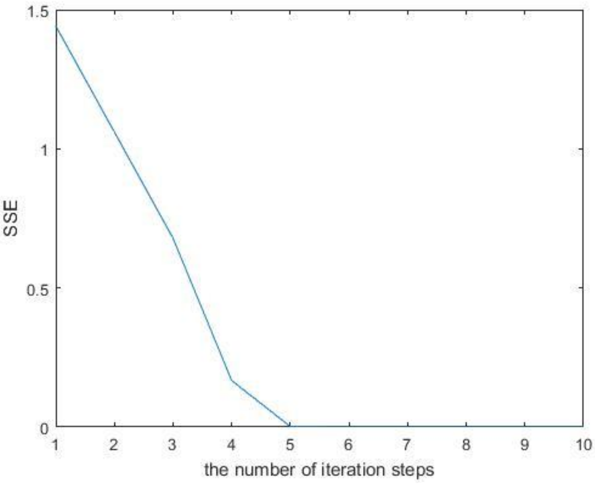

Fig. 1 Convergence of the c-FastICA algorithm

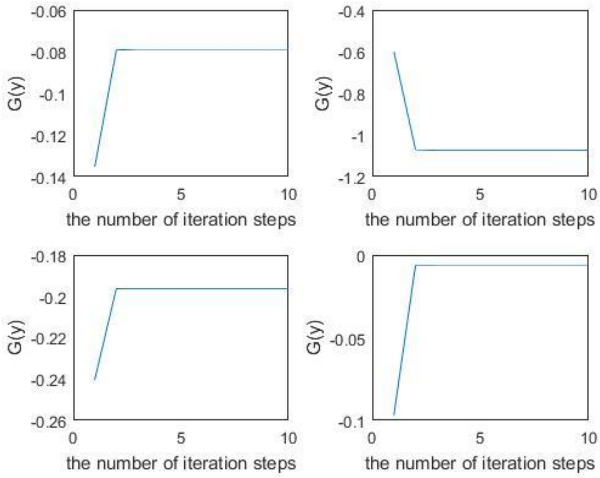

Fig. 2 Convergence of G for the c-FastICA algorithm

In this experiment, we choose the sum of squared deviation of $\boldsymbol{P}=|\boldsymbol{W}^H(\boldsymbol{Q}\boldsymbol{A})|$ from the nearest permutation matrix to measure effect of separation, which is because $|\boldsymbol{W}^H(\boldsymbol{Q}\boldsymbol{A})|$ should converge to a permutation matrix. The experimental results are shown in Fig. 1 and Fig. 2. Contrast function $G_2(y)=\log(a_2+y)$ is chosen, where $a_2=0.1$. And all the mathematical expectations in the test are replaced by sample mean values.From Fig. 1, we know algorithm convergence about needs five steps. Thus, Fig. 1 verifies the convergence of the c-FastICA algorithm. And Fig. 2 shows convergence of $G(\cdot)$. Obviously, $G(\cdot)$ can converge to local maxima or minima.

For the nc-FastICA algorithm, we produce noncircular source signals by $s_{nc} = real(s_j)\sin(f) + j * imag(s_j)\cos(f)$, $j = 1,2,3,4$, where f is used for controlling the degree of noncircularity with noncircularity index $\eta = \tan(f) = 40$. And the whitening step of the nc-FastICA algorithm is different from the one of the c-FastICA algorithm. For the nc-FastICA algorithm, the whitening step does not decorrelate the components as is done in the strong-uncorrelating transform (SUT). The specific whitening process can be found in [25]. The mixing matrix and related parameters of the nc-FastICA algorithm are also randomly generated by MATLAB. Iterative formula is as (21) for the nc-FastICA algorithm. In this experiment we choose contrast function as $G_2(y) = \log(a_2 + y)$, where $a_2 = 0.1$. The results of the experiment are shown in Fig. 3 and Fig. 4.

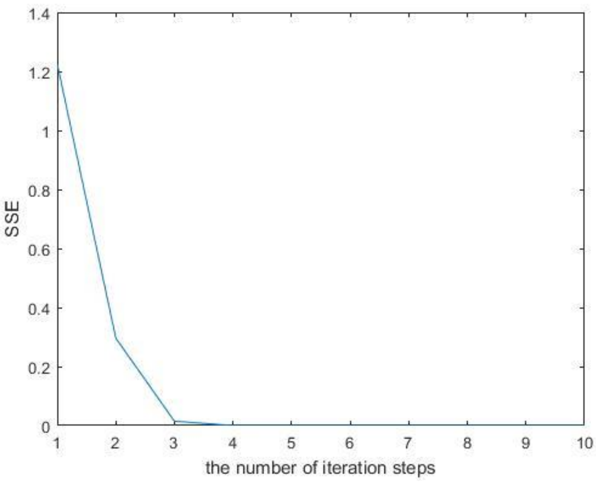

Fig. 3 Convergence of the nc-FastICA algorithm

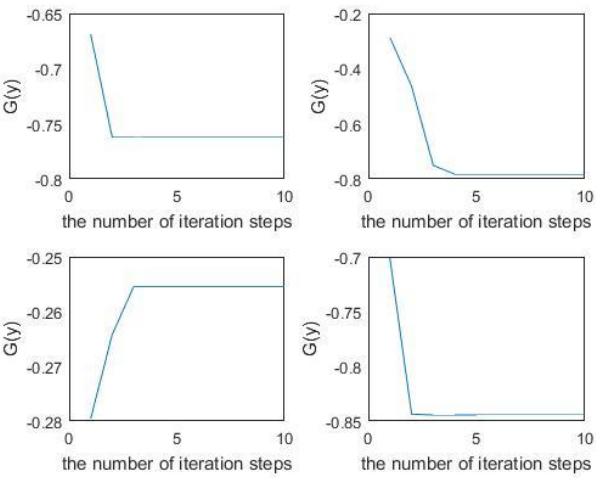

Fig. 4 Convergence of G for the nc-FastICA algorithm

From Fig. 3, we know algorithm convergence about needs four steps. Fig. 3 verifies the convergence of the nc-FastICA algorithm. And Fig. 4 shows convergence of $G(\cdot)$. Obviously, $G(\cdot)$ can converge to local maxima or minima.

The second experiment is conducted to test the relationship between performance and sample size for the c-FastICA algorithm and the nc-FastICA algorithm, respectively. Except taking the four source signals of the first experiment, this experiment adds another source signal with Beta Distribution, whose parameters are randomly generated. And the noncircularity index is chosen as 32. Similar to the performance index (PI) in real cases in [28], the performance of the different algorithms is measured by the normalized Amari index I_A [29]:

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

where p_{ij} $i = 1,\cdots,M$, $j = 1,\cdots,M$, are the entries of the matrix $\boldsymbol{P} = |\boldsymbol{W}^H(\boldsymbol{QA})|$. The lower the I_A value, the better the separation result. If $10\log I_A > -10\text{dB}$, the algorithm is not performing adequately. We test two algorithms with three classic nonlinearity as described in part III.

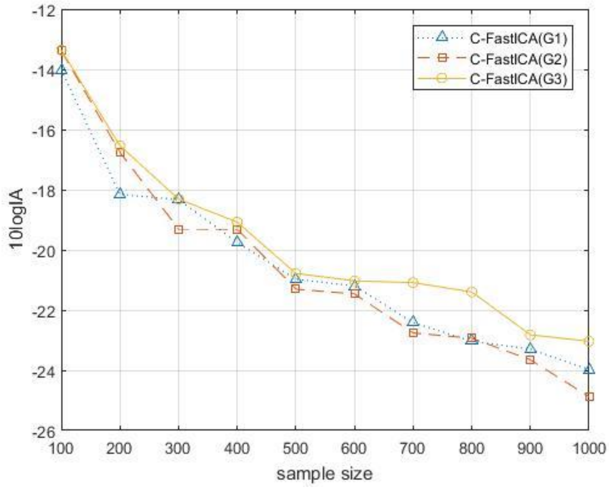

Fig. 5 The average $10\log I_A$ about sample size for the c-FastICA algorithm

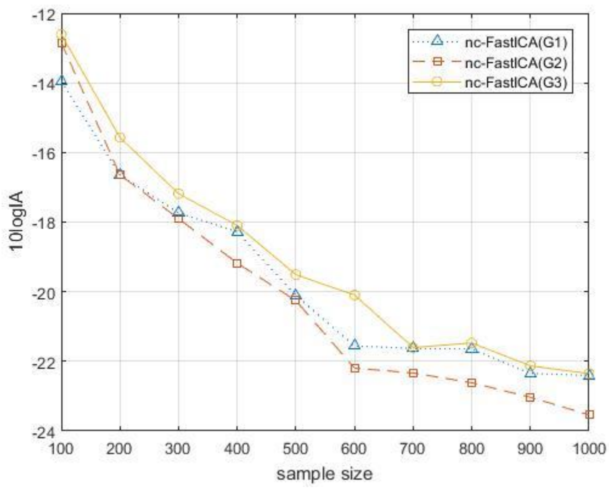

Fig. 6 The average $10\log I_A$ about sample size for the nc-FastICA algorithm

Fig. 5 shows the average $10\log I_A$ by varying sample size from 100 to 1000 for the c-FastICA algorithm. And Fig. 6 shows the average $10\log I_A$ for the nc-FastICA algorithm.

Analyzing Fig. 5 and Fig. 6, we can summary three properties of the c-FastICA algorithm and the nc-FastICA algorithm. (1) All of values $10\log I_A$ are less than 10. Thus, both algorithms perform well. (2) Six curves of average $10\log I_A$ decrease as sample size increases. Therefore, we draw a conclusion that the larger the sample size, the better the separation result for both the c-FastICA algorithm and the nc-FastICA algorithm. (3) Comparing separation effects of three nonlinearity , we can know that both two algorithms with nonlinearity G_3 perform worst. Separation effect of the c-FastICA algorithm with nonlinearity G_2 is best, when sample size is about larger than 450. And separation effect of the nc-FastICA algorithm with nonlinearity G_2 is best, when sample size is larger than 300. Thus, we usually choose G_2 as nonlinearity to solve practical problem.

The third experiment is conducted to verify poor separation performance for the c-FastICA algorithm and the nc-FastICA algorithm, when gaussian distribution is chosen. We generate circular complex random variables from Generalized Gaussian Distribution (GGD) [30], with probability density function defined by

$$p(b,p)=\frac{p}{b^{2/p}\Gamma(\frac{2}{p})2\pi}\exp(\frac{-|r|^p}{b}), \quad (31)$$

where b is the scale parameter and p is the shape parameter. $\Gamma(\cdot)$ is gamma function in complex field. By adjusting p, we can generate super-Gaussian variables with $0<p<2$, Gaussian variables with $p=2$, sub-Gaussian variables with $p>2$. Specially, we can generate Laplace distribution when $p=1$. In this experiment three cases with $p=1.25$, $p=2$, and $p=3$ are discussed. Variables r is generated as follows,

$$r=\text{gamrnd}(2/p,a)^{1/p}\exp(2\pi j*\text{rand}),$$

where $\text{gamrnd}(2/p,a)$ generates gamma random variables, rand generate uniformly distributed variables in $[0,1]$, and a is also randomly generated. We generate two circular complex source signals s_c by the above way. Then we produce two noncircular complex source signals s_{nc} by $s_{nc}=real(s_c)\sin(f)+j*imag(s_c)\cos(f)$ with the noncircularity index $\eta=\tan(f)=32$. Therefore, there are four source signals and two signals is noncircular signals with same noncircularity index. In this case, we only use nc-FastICA algorithm to separate signal. Sample size is chosen as 1000 in this part. The simulation results are shown in Table II. As shown in Table II, all the cases perform well when $p=1.2$ and $p=3$. However, algorithm performs poor when $p=2$. Thus, we verify poor separation performance, when gaussian distribution is chosen.

TABLE II
Average Separation Results for the Complex GGD Sources

I_A	$p=1.25$	$p=2$	$p=3$
nc-FastICA(G_1)	-22.0603	-9.05414	-22.3591
nc-FastICA(G_2)	-22.9127	-9.40878	-22.668
nc-FastICA(G_3)	-22.4322	-10.3976	-22.8133

VI. CONCLUSION

In this paper, the derivation of the nc-FastICA algorithm proposed by Novey M, Adali T is improved by using generalized linear (or linear-conjugation-linear) transformation. Further we get the c-FastICA algorithm. The theoretical basis of the proof is stronger by improving. Another main result of the article is proof about the relationship between the fixed point (pseudo-fixed point) of the iterative function and the local minima(maxima) point of contrast function. There are specific content and proof in theorem 1 and theorem 2. Theorem 3 applies this result to the c-FastICA algorithm. And we prove it by orthogonal projection.

REFERENCES

[1] P. Comon, "Independent component analysis — a new concept?" *Signal Processing*, vol.36, pp. 287–314, 1944.

[2] A. Hyvärinen, J. Karhunen, and E. Oja, "Independent Component Analysis," *New York: Wiley*, 2001.

[3] C. Andrzej, A. Shunichi, J. Wiley, "Adaptive blind signal and image processing learning algorithms and applications," *IEEE Transactions on Neural Networks*, vol. 14, no. 6, pp. 1580-1580, 2004.

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

[5] A. Hyvärinen, E. Oja, "A Fast Fixed-Point Algorithm for Independent Component Analysis," *International Journal of Neural Systems*, vol. 10, no. 01, pp. 1-8, 2000.

[6] E. Ollila, "The Deflation-Based FastICA Estimator: Statistical Analysis Revisited," *IEEE Transactions on Signal Processing*, vol. 58, no. 3, pp. 1527-1541, 2010.

[7] E. Oja, "Convergence of the symmetrical FastICA algorithm," *International Conference on Neural Information Processing*, 2002.

[8] E. Bingham, A. Hyvärinen, "A fast fixed-point algorithm for independent component analysis of complex valued signals," *International journal of neural systems*, vol. 10, no. 01, pp. 1-8, 2000.

[9] J. F. Cardoso, A. Souloumiac, "An efficient technique for the blind separation of complex sources," *IEEE Signal Processing Workshop on Higher-order Statistics*. IEEE, 1993.

[10] S. Fiori, "Extended hebbian learning for blind separation of complex-valued sources," *IEEE Transactions on Circuits and Systems II: Analog and Digital Signal Processing*, vol. 50, no. 4, pp. 195-202, 2003.

[11] M. Novey, T. Adali, "On Extending the Complex FastICA Algorithm to Noncircular Sources," *IEEE Transactions on Signal Processing*, vol. 56, no. 5, pp. 2148-2154, 2008.

[12] X. L. Li, T. Adali, "Complex Independent Component Analysis by Entropy Bound Minimization," *IEEE Transactions on Signal Processing*, vol. 58, no. 10, pp. 5151-5164, 2010.

[13] B. Loesch, B. Yang, "Cramér-Rao bound for circular and noncircular complex independent component analysis," *IEEE Transactions on Signal Processing*, vol. 61, no. 2, pp. 365-379, 2013.

[14] J. C. Chao, S. C. Douglas, "A robust complex FastICA algorithm using the huber M-estimator cost function," *International Conference on Independent Component Analysis & Signal Separation*, 2007.

[15] G. Qian, W. Ping, H. Liao, "Efficient Variant of Noncircular Complex FastICA Algorithm for the Blind Source Separation of Digital Communication Signals," *Circuits Systems & Signal Processing*, vol. 35, no. 2, pp. 705-717, 2016.

[16] Z. Ruan, L. Li, G. Qian, "On extending the complex FastICA algorithms to noisy data," *Neural Networks*, vol. 60, no. C, pp. 194-202, 2014.

[17] G. Qian, L. Li, H. Liao, "Performance Analysis of the nc-FastICA Algorithm," *Circuits Systems & Signal Processing*, vol. 34, no.2, pp. 441-457, 2015.

[18] G. S. Fu, R. Phlypo, M. Anderson, *et al*, "Complex Independent Component Analysis Using Three Types of Diversity: Non-Gaussianity, Nonwhiteness, and Noncircularity," *IEEE Transactions on Signal Processing*, vol. 63, no. 3, pp. 794-805, 2015.

[19] W. Zhao, Y. Shen, P. Xu, *et al*, "Efficient Optimization of Reference-Based Negentropy for Noncircular Sources in Complex

ICA," *Circuits Systems and Signal Processing*, vol. 35, no. 12, pp. 1-23, 2016.
[20] Mikhael, B. Wasfy, Ranganathan, *et al*, "Complex Adaptive ICA Employing the Conjugate Gradient Technique for Signal Separation in Time-Varying Flat Fading Channels," *Circuits Systems & Signal Processing*, vol. 29, no. 3, pp. 469-480, 2010.
[21] Z. Yang, S. A. Kassam, "Optimum nonlinearity and approximation in complex FastICA," *Information Sciences & Systems*, 2012.
[22] P. J. Schreier, L. L. Scharf, "Statistical Signal Processing of Complex-Valued Data: The Theory of Improper and Noncircular Signals," *Cambridge: Cambridge University Press*, 2010.
[23] E. Oja, Z. Yuan, "The FastICA Algorithm Revisited: Convergence Analysis," *IEEE Transactions on Neural Networks*, vol. 17, no. 6, pp. 1370-1381, 2006.
[24] H. Shen, M. Kleinsteuber, K. Huper, "Local Convergence Analysis of FastICA and Related Algorithms," *IEEE Transactions on Neural Networks*, vol. 19, no. 6, pp. 1022-1032, 2008.
[25] J. Eriksson, V. Koivunen, "Complex-valued ICA using second order statistics," *Machine Learning for Signal Processing, IEEE Signal Processing Society Workshop*, 2004.
[26] W. Wirtinger, "On the formal theory of functions of more complex variables," *Mathematical Annals*, vol. 97, no. 1, pp. 357-375, 1927. "Zur formalen Theorie der Funktionen von mehr komplexen Veränderlichen," *Mathematische Annalen*, vol. 97, no. 1, pp. 357-375, 1927.
[27] D. B. A. Van, "Complex gradient and Hessian," *IEEE Processing visualization image Signal Process*, vol. 141, no. 6, pp. 380-382, 1994.
[28] J. Ye, H. Jin, Q. Zhang, "Adaptive weighted orthogonal constrained algorithm for blind source separation ☆," *Digital dignal Processing*, vol. 23, no. 2, pp. 514-521, 2013.
[29] S. Amari, A. Cichocki, H. H. Yang, "A new learning algorithm for blind signal separation," *In Advances in Neural Information Processing Systems*, vol. 8, pp. 757–763, 1996.
[30] E. W. Stacy, "A Generalization of the Gamma Distribution," *The Annals of Mathematical Statistics*, vol. 33, no. 3, pp. 1187-1192, 1962.

Qianru Ma received her B.S. degree from Anshan normal college. She is currently a master student at Xidian University. Her scientific interest is blind source separation using independent component analysis.

Jimin Ye received the B.S., M.S. degrees from Shaanxi Normal University. And he received Ph.D. degree from Xidian University. He is currently a professor in Xidian University, Xi'an China. His scientific interests are risk theory, statistical signal processing, and blind source separation using independent component analysis.