Hindawi Publishing Corporation
EURASIP Journal on Advances in Signal Processing
Volume 2007, Article ID 36525, 15 pages
doi:10.1155/2007/36525

Research Article

Fixed-Point Algorithms for the Blind Separation of Arbitrary Complex-Valued Non-Gaussian Signal Mixtures

Scott C. Douglas

Department of Electrical Engineering, School of Engineering, Southern Methodist University, P.O. Box 750338, Dallas, newline TX 75275, USA

Received 1 October 2005; Revised 10 May 2006; Accepted 22 June 2006

Recommended by Andrzej Cichocki

We derive new fixed-point algorithms for the blind separation of complex-valued mixtures of independent, noncircularly symmetric, and non-Gaussian source signals. Leveraging recently developed results on the separability of complex-valued signal mixtures, we systematically construct iterative procedures on a kurtosis-based contrast whose evolutionary characteristics are identical to those of the FastICA algorithm of Hyvarinen and Oja in the real-valued mixture case. Thus, our methods inherit the fast convergence properties, computational simplicity, and ease of use of the FastICA algorithm while at the same time extending this class of techniques to complex signal mixtures. For extracting multiple sources, symmetric and asymmetric signal deflation procedures can be employed. Simulations for both noiseless and noisy mixtures indicate that the proposed algorithms have superior finite-sample performance in data-starved scenarios as compared to existing complex ICA methods while performing about as well as the best of these techniques for larger data-record lengths.

Copyright © 2007 Hindawi Publishing Corporation. All rights reserved.

1. INTRODUCTION

Both blind source separation (BSS) and independent component analysis (ICA) are concerned with m-dimensional linear signal mixtures of the form

$$\mathbf{x}(k) = \mathbf{A}\mathbf{s}(k), \tag{1}$$

where $\mathbf{A}$ is an unknown $(m \times m)$ mixing matrix and $\mathbf{s}(k) = [s_1(k) \ \cdots \ s_m(k)]^T$ is a vector-valued signal of sources. In most treatments of either task in the scientific literature, the sources $\{s_i(k)\}$ are assumed to be statistically independent and real-valued, and the matrix $\mathbf{A}$ is assumed to be full rank. If certain additional separability conditions are met, it is possible to compute a demixing matrix $\mathbf{B}$ such that

$$\mathbf{y}(k) = \mathbf{B}\mathbf{s}(k) \tag{2}$$

contains independent elements that are possibly scaled and shuffled with respect to the sources in $\mathbf{s}(k)$. Separation or extraction of the independent components is considered successful in such cases, as demixing of the mixed sources has been achieved. Numerous algorithms have been developed for separating real-valued mixtures, including maximum-likelihood information-theoretic approaches [1–4], contrast-based approaches [5–7], and decorrelation-based approaches [8–10]. Among these methods, the FastICA procedure in [7] has a number of nice features, including fast convergence, global convergence for kurtosis-based contrasts, and the lack of any step-size parameter. For a kurtosis-based measure of negentropy, the FastICA algorithm employs a separation criterion similar to other approaches involving cumulant-based contrasts [5, 6], although the optimization method employed by the FastICA algorithm is quite different from the joint diagonalization procedures employed in other approaches.

Consider now the case where $\mathbf{A}$ and $\mathbf{s}(k)$ are complex-valued, such that $\mathbf{A} = \mathbf{A}_R + j\mathbf{A}_I$, $\mathbf{s}(k) = \mathbf{s}_R(k) + j\mathbf{s}_I(k)$, and $s_i(k) = s_{R,i}(k) + js_{I,i}(k)$, where $j = \sqrt{-1}$. Separating complex (-valued) linear signal mixtures is important for a number of tasks of practical interest, such as in cochannel interference mitigation for wireless communications and array processing applications and in the decomposition of biomedical imagery for medical diagnosis [11–14]. Fewer algorithms for separating complex signal mixtures have been described in the scientific literature. Examples of such algorithms include JADE [5], a complex-valued extension of the FastICA algorithm [15], and maximum-likelihood approaches [11, 13]. In [15], the complex-valued source signals have been assumed to be *circular*, such that the probability

density function (p.d.f.) of $s_i(k)$ depends only on its modulus $|s_i(k)| = \sqrt{s_R^2(k) + s_I^2(k)}$, a restrictive assumption.

Recently, it has been shown that complex ICA has a specific statistical and mathematical structure that is distinct from the real-valued case [16–18]. In particular, it is possible to identify the matrix $\mathbf{A}$ up to scaling and permutation factors in cases where $\mathbf{s}(k)$ contains multiple complex noncircular Gaussian-distributed sources, a situation distinct from the real-valued case. The key concept behind these novel results is the relaxing of the circularity assumptions of the distributions of the complex sources $\{s_i(k)\}$, such that each $s_i(k)$ has a generic but unstructured p.d.f. $p_i(s_i) = p_i(s_{R,i}, s_{I,i})$. Algorithms for separating mixtures of such general-form complex sources have appeared only recently [19, 20], and extensions of the most popular algorithms have yet to be considered.

In this paper, we present a careful study of the complex-valued ICA and BSS tasks for non-Gaussian signal mixtures. Both noncircular and circular independent source signals are considered. The role of decorrelation in complex-valued ICA is carefully delineated, where the results of [18] are taken into account. We then present several extensions of the popular FastICA algorithm for fourth-moment separation criteria to the noncircular complex-valued case. Unlike the derivation in [15], our approach to constructing the algorithms exploits the structure of the fourth-moment symmetric tensor of the source signal vector to generate an update relation that preserves the fast and efficient convergence properties of the fixed-point iteration[1] as obtained by the original FastICA algorithm for a kurtosis contrast in the real-valued case [7]. Our various algorithms differ in the way they treat the real and imaginary portions of the sources $\{s_i(k)\}$ depending on whether or not $s_R(k)$ and $s_I(k)$ are statistically independent. Brief convergence proofs of the algorithms are given showing that they achieve separation in the case where $\mathbf{s}(k)$ contains at least $(m-1)$ non-Gaussian-distributed sources. Simulations are then provided to indicate their separating capabilities for complex-valued BSS tasks.

2. ON COMPLEX-VALUED RANDOM VARIABLES

Because our work focuses on the separation of a general class of complex-valued signal mixtures, it is important to delineate the statistical structure of these sources. We will later use the described statistical structure to develop efficient separation algorithms for noncircular sources.

Let $s(k) = s_R(k) + js_I(k)$ denote a scalar complex-valued random variable with p.d.f. $p(s_R, s_I)$. The marginal p.d.f.'s of $s_R(k)$ and $s_I(k)$ are

$$\begin{aligned} p_R(s_R) &= \int_{-\infty}^{\infty} p(s_R, s_I)\, ds_I, \\ p_I(s_I) &= \int_{-\infty}^{\infty} p(s_R, s_I)\, ds_R, \end{aligned} \tag{3}$$

respectively. Let $g(s(k)) = g_R(s_R(k), s_I(k)) + jg_I(s_R(k), s_I(k))$ be an arbitrary complex function of $s(k)$, and define the expectation operator as

$$E\{g(s(k))\} = \iint_{-\infty}^{\infty} [g_R(s_R, s_I) + jg_I(s_R, s_I)]\, p(s_R, s_I)\, ds_R\, ds_I. \tag{4}$$

For convenience, we will assume that $s(k)$ is a zero-mean random variable, such that $E\{s(k)\} = E\{s_R(k)\} = E\{s_I(k)\} = 0$. The complex conjugate of $s(k)$ is denoted as $s^*(k) = s_R(k) - js_I(k)$.

Let $y(k) = cs(k)$, where $c = c_R + jc_I$ is a complex scalar. Clearly, $E\{y(k)\} = E\{y_R(k)\} = E\{y_I(k)\} = 0$ for any complex scalar c. Then, the following theorem relates to the distribution of $y(k)$, the proof of which is in Appendix A.

Theorem 1. *For any zero-mean complex r.v. $s(k)$ satisfying $E\{s_R^2(k)\} < \infty$ and $E\{s_I^2(k)\} < \infty$, it is always possible to find a complex scalar c such that $y(k)$ has the following properties:*

$$E\{|y(k)|^2\} = 1, \tag{5}$$

$$E\{[y(k)]^2\} = \lambda, \tag{6}$$

where λ is a real number satisfying $0 \le \lambda \le 1$.

Corollary 1. *Under such scaling, the random variable $y(k)$ has the following additional properties:*

$$\begin{aligned} E\{(y_R(k))^2\} &= \frac{1+\lambda}{2}, \\ E\{(y_I(k))^2\} &= \frac{1-\lambda}{2}, \\ E\{y_R(k)y_I(k)\} &= 0. \end{aligned} \tag{7}$$

Corollary 2. *The power of $y_R(k)$ is greater than or equal to that of $y_I(k)$ with equality if and only if $E\{[y(k)]^2\} = 0$.*

The above theorem and corollaries show that it is always possible to "scale" a complex-valued random variable so that (a) its power is unity, (b) the power of its imaginary part is not greater than that of its real part, and (c) its real and imaginary parts are uncorrelated. Such signals are said to be *strong-uncorrelated*, in deference to the terminology developed in [18]. For this reason, we will in the sequel assume that $s_i(k)$ possesses this statistical structure, as we can always absorb the complex scaling factor c for each source into the mixing matrix $\mathbf{A}$ within the model in (1). Note that this structure says nothing about the independence of $s_R(k)$ and $s_I(k)$ (e.g., they can be statistically dependent) or about the distribution of $s_i(k)$ (e.g., it can be non-Gaussian).

It should also be noted that if $s(k)$ is circular such that $p(s) = \overline{p}(|s|)$, then $E\{s_R(k)s_I(k)\} = 0$, such that any complex-valued scalar c satisfying $|c|^2 = 1/E|s(k)|^2$ satisfies the conditions in (7). In such cases, $\lambda = 0$. The condition $E\{s_R(k)s_I(k)\} = 0$ does not guarantee circularity; however, a good practical example is the family of discrete-valued constant-modulus sources that includes 4QAM and 8-PSK whose distributions depend on the angle of $s(k)$.

[1] Technically, the FastICA algorithm attempts to find coefficient vectors that point in a *fixed direction* but may oscillate back in forth in absolute sign. For historical reasons, we adopt the same terminology in [7] for this class of algorithms.

This paper will be concerned with algorithms that exploit the fourth-order moment structure of the vector $\mathbf{s}(k)$. Fourth-order cumulants have been heavily exploited in the development of ICA, BSS, and blind deconvolution approaches in the real-valued case, so it is reasonable to consider their structure in developing separation algorithms for the complex case. The following theorem and associated corollaries give the fourth-order moment properties of i.i.d. sources $\{s_i(k)\}$ that are strong-uncorrelated. Proofs are again given in Appendix B.

Theorem 2. *Assume that* $\mathbf{s}(k)$ *contains* m *zero-mean, independent, strong-uncorrelated signals* $s_i(k)$, $1 \le i \le m$, *where* $E\{|s_i(k)|^2\} = 1$ *and* $E\{s_i^2(k)\} = \lambda_i$, $0 \le \lambda_i \le 1$. *Define the symmetric fourth-order moment tensor*

$$K_{ijln} = E\{s_i(k)s_j^*(k)s_l^*(k)s_n(k)\}. \tag{8}$$

Then, the values of K_{ijln} *are*

$$K_{ijln} = \begin{cases} 1 & \text{if } i = j \neq l = n \text{ or } i = l \neq j = n, \\ \lambda_i\lambda_j & \text{if } i = n \neq j = l, \\ \kappa_i + 2 + \lambda_i^2 & \text{if } i = j = l = n, \\ 0 & \text{otherwise}, \end{cases} \tag{9}$$

where κ_i *is the symmetric kurtosis defined as*

$$\kappa_i = E\{|s_i(k)|^4\} - 2(E\{|s_i(k)|^2\})^2 - |E\{s_i^2\}|^2 \tag{10}$$

$$= E\{|s_i(k)|^4\} - 2 - \lambda_i^2. \tag{11}$$

Corollary 3. *Let* $s_i(k)$ *be a strong-uncorrelated Gaussian r.v. with distribution*

$$p_G(s_R, s_I) = \frac{1}{\pi\sqrt{1-\lambda^2}} \exp\left(-\left[\frac{s_R^2}{(1+\lambda)} + \frac{s_I^2}{(1-\lambda)}\right]\right), \tag{12}$$

where $0 \le \lambda \le 1$. *Then, the symmetric kurtosis of* $s_i(k)$ *is zero.*

Because of the importance of the kurtosis in our derivations, we will define the kurtosis operator for a complex random variable $s(k)$ as

$$\kappa[s(k)] = E\{|s(k)|^4\} - 2(E\{|s_i(k)|^2\})^2 - |E\{s_i^2\}|^2, \tag{13}$$

where $\kappa[s_i(k)] = \kappa_i$.

The symmetric fourth-order moment tensor K_{ijln} for independent and strong-uncorrelated complex random vectors is similar in structure to that of independent real-valued random vectors, in which $\lambda = 1$, and independent circularly complex random vectors, in which $\lambda = 0$. In particular, terms that depend on the third-order moments vanish in all three cases. For independent $\{s_i(k)\}$ in the noncircular complex case, however, only *independent and strong-uncorrelated* random variables maintain this nice structure. This fact underscores the importance of transformations that impose a strong-uncorrelated structure to a random vector, a fact that will play an important role when we develop algorithms for separating non-Gaussian complex sources in the following sections.

3. ON THE EXTRACTION OF A SINGLE COMPLEX-VALUED SOURCE

Consider an algorithm that adjusts a single row of the separation matrix $\mathbf{B}$ in an attempt to extract a single source $s_i(k)$. Let $\mathbf{b} = [b_1 \ \cdots \ b_m]^T$ denote the transposed version of this row vector. Define the output signal at time k as

$$y(k) = \mathbf{b}^T\mathbf{x}(k). \tag{14}$$

Assuming that $\mathbf{A}$ is full rank, we can write the output signal in terms of the combined coefficient vector $\mathbf{c}$ given by

$$\mathbf{c} = \mathbf{A}^T\mathbf{b}, \tag{15}$$

in which case

$$y(k) = \mathbf{c}^T\mathbf{s}(k). \tag{16}$$

Then, the following theorem and corollary relate to the moments of $y(k)$, the proofs of which are in Appendix C.

Theorem 3. *For a source vector that contains independent, zero-mean, possibly noncircular, and strong-uncorrelated sources* $\{s_i(k)\}$, *the output signal* $y(k)$ *has the following moments:*

$$E\{y(k)\} = 0, \tag{17}$$

$$E\{|y(k)|^2\} = \sum_{i=1}^{m} |c_i|^2, \tag{18}$$

$$E\{[y(k)]^2\} = \sum_{i=1}^{m} \lambda_i c_i^2, \tag{19}$$

$$E\{|y(k)|^4\} = \sum_{i=1}^{m} \kappa_i |c_i|^4 + 2\left(\sum_{i=1}^{m} |c_i|^2\right)^2 + \left(\sum_{i=1}^{m} \lambda_i c_i^2\right)^2. \tag{20}$$

Corollary 4. *The kurtosis of* $y(k)$ *is*

$$\kappa[y(k)] = \sum_{i=1}^{m} \kappa_i |c_i|^4. \tag{21}$$

The result in (21) indicates two important facts in separating mixtures of noncircular complex-valued independent sources.

(i) The kurtosis of $y(k)$ as represented in the combined coefficient space depends on the circularity coefficients $\{\lambda_i\}$ of the noncircular sources only through the values κ_i in (11) for strong-uncorrelated sources.

(ii) Consider the representation of each c_i in complex polar form as

$$c_i = A_i e^{j\theta_i}. \tag{22}$$

Then, the kurtosis of $y(k)$ only depends on the amplitudes $\{A_i\}$ of the coefficients in the combined coefficient space and is independent of the complex phases

of these coefficients. Moreover, through this polar representation, we can represent the kurtosis and power of $y(k)$ as

$$\kappa[y(k)] = \sum_{i=1}^{m} \kappa_i A_i^4, \tag{23}$$

$$E\{|y(k)|^2\} = \sum_{i=1}^{m} A_i^2. \tag{24}$$

Equations (23)-(24) have appeared before in the contexts of single-channel blind deconvolution for filtered complex-valued sequences (cf. [21]) and of blind source separation for real-valued signal mixtures (cf. [6, 22, 23]). In blind deconvolution tasks, there is only one kurtosis value $\kappa_i = \kappa$ in (23), which simplifies the optimization strategy for achieving a deconvolved sequence. In real-valued blind source separation, the real-valued combined system coefficients play roles that are identical to those of the amplitudes of the combined system coefficients in the complex-valued case. It is this latter correspondence that allows us to directly state an optimization strategy for extracting a single complex-valued source, as indicated in the following theorem.

Theorem 4. *Consider the single-unit extraction criterion*

$$\mathcal{J}(\mathbf{b}) = \left| \frac{\kappa[y(k)]}{\left(E\{|y(k)|^2\}\right)^2} \right|, \tag{25}$$

where $y(k) = \mathbf{b}^T\mathbf{x}(k)$. Assume that at least one of the sources has a nonzero kurtosis $\kappa_i \neq 0$. Then, maximization of $\mathcal{J}(\mathbf{b})$ over all possible $\mathbf{b}$ under the constraint that $E\{|y(k)|^2\} = 1$ yields one of the columns of $\mathbf{A}^{-1}$ for which $\kappa_i \neq 0$ up to a complex unit-modulus scaling factor.

Proof. As stated previously, the relations in (23)-(24) are identical in form to those in the real-valued blind source separation case, where the roles of the real-valued amplitudes $\{A_i\}$ in the complex-valued separation case play identical roles to those of the real-valued combined system coefficients $\{c_i\}$ in the real-valued separation case. Thus, we directly borrow from existing proofs in the literature, such as [22], where it has already been shown that maximization of $\mathcal{J}(\mathbf{b})$ under unit-output-power constraints occurs only at points corresponding to an extracted source, such that A_i is nonzero for a single index $i \in \{1, \leq, m\}$. The constraint $A_i = 1$ then follows from the unit-power constraint and (24). In practical implementations, prewhitening is employed to translate this unit-power constraint to a unit-norm coefficient constraint. □

4. FIXED-POINT ALGORITHMS FOR EXTRACTING A SINGLE ARBITRARY COMPLEX SOURCE

4.1. Preliminaries

Blind source separation requires the extraction of all m sources in the linear mixture $\mathbf{x}(k)$. The FastICA algorithm with generalized contrast locally maximizes a chosen cost function to achieve separation. For real-valued signal mixtures, the FastICA algorithm that maximizes absolute values of signal kurtoses is a simple and efficient separation technique. It is fast, globally convergent, devoid of any step size parameters, and will extract all sources in the mixture as long as all but one of their kurtosis values are nonzero. For these reasons, we now explore extensions of the FastICA algorithm with kurtosis contrast for separating mixtures of noncircular complex-valued independent sources.

In [7], the FastICA algorithm for real-valued mixtures is derived as an approximate Newton procedure for maximizing a set of continuous-valued generalized contrast functions. When the kurtosis is employed as a contrast, the algorithm has a particularly appealing form when expressed in the combined system coefficient vector $\mathbf{c}_t$ at iteration t, as shown in [7] (see also [24]):

$$\widetilde{\mathbf{c}}_t = \mathbf{K}\mathbf{F}(\mathbf{c}_t), \tag{26}$$

$$\mathbf{c}_{t+1} = \frac{\widetilde{\mathbf{c}}_t}{\sqrt{\widetilde{\mathbf{c}}_t^T\widetilde{\mathbf{c}}_t}}, \tag{27}$$

where $\mathbf{K}$ is a diagonal matrix of source kurtoses and $\mathbf{F}(\mathbf{c}_t)$ is a diagonal matrix whose ith diagonal entry is c_{it}^3. While the derivation of the FastICA algorithm in the real-valued case is theoretically appealing, the real utility of the FastICA procedure can be inferred from the form of (26)-(27), which leads to cubic convergence near a separating solution. Moreover, its average performance over a uniform prior of initial coefficient vector directions as the number of iterations increases becomes exponential with a rate of (1/3); see [24–26] for more discussion of these issues. For these reasons, in what follows we attempt to find an algorithm whose coefficient updates in the combined system coefficient vector $\mathbf{c}_t = \mathbf{A}^T\mathbf{b}_t$ obey a similar relation as (26)-(27) in the limit as the data-record length tends to infinity, where the amplitudes of the elements of $\mathbf{c}_t$ in the complex-valued case behave as the (absolute values of) the elements of $\mathbf{c}_t$ in the real-valued case. This method of derivation is an alternative to that using complex differentiation, which involves different rules depending on the choice of differentiation operator [18]. It leverages the main reason why the FastICA algorithm is so popular in ICA and blind source separation tasks: the underlying structure of (26)-(27) allows the algorithm to converge quickly, in a way that is largely independent of the distributions of the sources being extracted. As will be seen, the derivation of these algorithms for noncircular sources requires the careful expression and evaluation of the second-order noncircular statistical properties of the source signals in order to obtain convergent behavior similar to that in (26)-(27). The method described in [15] has unknown convergence performance when the sources are noncircular.

Our derivation assumes that we have a set of N measurements $\mathbf{x}(n)$, $1 \leq n \leq N$, from a complex mixture model of the form in (1), where

$$\begin{aligned} \frac{1}{N}\sum_{n=1}^{N} \mathbf{s}(n)\mathbf{s}^H(n) &= \mathbf{I} + \mathbf{\Delta}_R, \\ \frac{1}{N}\sum_{n=1}^{N} \mathbf{s}(n)\mathbf{s}^T(n) &= \mathbf{\Lambda} + \mathbf{\Delta}_P, \end{aligned} \tag{28}$$

where $\mathbf{\Delta}_R$ and $\mathbf{\Delta}_P$ are matrices of small Frobenius norm caused by finite-sample effects. The elements of $\mathbf{s}(n)$ are realizations of m statistically independent complex-valued random processes, in which at most one of these random processes has a zero kurtosis.

4.2. Algorithm based on the strong-uncorrelating transform

Our first fixed-point algorithm for noncircular complex-valued sources will rely on the strong-uncorrelating transform for signal prewhitening. The strong-uncorrelating transform as defined in [17] is a transformation that diagonalizes both the covariance matrix and pseudocovariance matrix given by

$$\begin{aligned}\mathbf{R}_{XX} &= \frac{1}{N}\sum_{n=1}^{N}\mathbf{x}(n)\mathbf{x}^H(n),\\ \mathbf{P}_{XX} &= \frac{1}{N}\sum_{n=1}^{N}\mathbf{x}(n)\mathbf{x}^T(n),\end{aligned} \tag{29}$$

respectively. For noncircular sources, the pseudocovariance matrix $\mathbf{P}_{XX}$ is nonzero. The strong-uncorrelating transform is defined by a matrix $\mathbf{G}$ such that

$$\begin{aligned}\mathbf{G}\mathbf{R}_{XX}\mathbf{G}^H &= \mathbf{I},\\ \mathbf{G}\mathbf{P}_{XX}\mathbf{G}^T &= \hat{\mathbf{\Lambda}},\end{aligned} \tag{30}$$

where $\hat{\mathbf{\Lambda}}$ is a diagonal real-valued matrix of ordered diagonal entries $1 \geq \hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \cdots \geq \hat{\lambda}_m \geq 0$. It is always possible to find a $\mathbf{G}$ such that (30) is satisfied. Methods for computing the strong-uncorrelating transform are given in [17, 18]. With this transformation, define the prewhitened signal vector

$$\mathbf{v}(k) = \mathbf{G}\mathbf{x}(k), \tag{31}$$

such that

$$\begin{aligned}\mathbf{R}_{VV} &= \frac{1}{N}\sum_{n=1}^{N}\mathbf{v}(n)\mathbf{v}^H(n) = \mathbf{I},\\ \mathbf{P}_{VV} &= \frac{1}{N}\sum_{n=1}^{N}\mathbf{v}(n)\mathbf{v}^T(n) = \hat{\mathbf{\Lambda}}.\end{aligned} \tag{32}$$

Under prewhitening, the relationship between $v(k)$ and $\mathbf{s}(k)$ is

$$\mathbf{v}(k) = \mathbf{\Gamma}\mathbf{s}(k), \tag{33}$$

where $\mathbf{\Gamma}$ is Hermitian ($\mathbf{\Gamma}\mathbf{\Gamma}^H = \mathbf{\Gamma}^T\mathbf{\Gamma}^* = \mathbf{I}$). The matrix $\mathbf{\Gamma}$ also obeys the property[2]

$$\mathbf{\Gamma}\left[\frac{1}{N}\sum_{n=1}^{N}\mathbf{s}(n)\mathbf{s}^T(n)\right]\mathbf{\Gamma}^T = \hat{\mathbf{\Lambda}}. \tag{34}$$

[2] If the sample pseudocovariance matrix of $\mathbf{s}(k)$ is exactly diagonal, then $\hat{\mathbf{\Lambda}} = \mathbf{\Lambda}$. Moreover, if the sample pseudocovariance matrix of $\mathbf{s}(k)$ is exactly diagonal with distinct positive entries, then $\hat{\mathbf{\Lambda}} = \mathbf{I}$ and $\mathbf{G} = \mathbf{A}^{-1}$. It should be noted, however, that $\hat{\mathbf{\Lambda}}$ is still diagonal even under finite-sample effects.

Consider first a single-source extraction task, in which

$$y(k) = \mathbf{w}^T\mathbf{v}(k), \tag{35}$$

where $\mathbf{w}$ is an m-dimensional vector of parameters to be adjusted. The relationship between $\mathbf{w}$ and the combined system coefficient vector is

$$\mathbf{c} = \mathbf{\Gamma}^T\mathbf{w}. \tag{36}$$

The second moment of the output signal is

$$\frac{1}{N}\sum_{n=1}^{N}|y(n)|^2 = \mathbf{w}^T\mathbf{R}_{VV}\mathbf{w}^* = \|\mathbf{w}\|^2 = \|\mathbf{c}\|^2 \tag{37}$$

and the fourth moment of the output signal can be written as

$$\begin{aligned}&\frac{1}{N}\sum_{n=1}^{N}|y(n)|^4\\ &\quad= \mathbf{w}^T\left[\frac{1}{N}\sum_{n=1}^{N}\mathbf{v}(n)\mathbf{v}^H(n)\mathbf{w}^*\mathbf{w}^T\mathbf{v}(n)\mathbf{v}^H(n)\right]\mathbf{w}^*\\ &\quad= \mathbf{w}^T\mathbf{\Gamma}\left[\frac{1}{N}\sum_{n=1}^{N}\mathbf{s}(n)\mathbf{s}^H(n)\mathbf{\Gamma}^H\mathbf{w}^*\mathbf{w}^T\mathbf{\Gamma}\mathbf{s}(n)\mathbf{s}^H(n)\right]\mathbf{\Gamma}^H\mathbf{w}^*\\ &\quad= \mathbf{c}^T\widehat{\mathbf{M}}\mathbf{c}^* = \mathbf{c}^H\widehat{\mathbf{M}}^T\mathbf{c},\end{aligned} \tag{38}$$

where we have defined the matrix $\widehat{\mathbf{M}}$ as

$$\widehat{\mathbf{M}} = \frac{1}{N}\sum_{n=1}^{N}\mathbf{s}(n)\mathbf{s}^H(n)\mathbf{c}^*\mathbf{c}^T\mathbf{s}(n)\mathbf{s}^H(n). \tag{39}$$

The following theorem gives the structure of $\widehat{\mathbf{M}}$, the proof of which is in Appendix D.

Theorem 5. *In the limit as $N \to \infty$, the value of $\widehat{\mathbf{M}}$ becomes*

$$\lim_{N\to\infty}\widehat{\mathbf{M}} = \mathbf{M} = \mathbf{c}^*\mathbf{c}^T + \mathbf{I}\mathbf{c}^H\mathbf{c} + \mathbf{\Lambda}\mathbf{c}\mathbf{c}^H\mathbf{\Lambda} + \mathbf{K}\,\mathrm{diag}\{\mathbf{c}\mathbf{c}^H\}, \tag{40}$$

where $\mathrm{diag}\{\mathbf{c}\mathbf{c}^H\}$ *is a diagonal matrix whose diagonal entries are the diagonal elements of the matrix* $\mathbf{c}\mathbf{c}^H$.

Using this result, we can approximate

$$\widehat{\mathbf{M}}^T\mathbf{c} \approx \mathbf{K}\,\mathrm{diag}\{\mathbf{c}\mathbf{c}^H\}\mathbf{c} + \mathbf{c}[2\mathbf{c}^H\mathbf{c}] + \mathbf{\Lambda}\mathbf{c}^*[\mathbf{c}^T\mathbf{\Lambda}\mathbf{c}]. \tag{41}$$

As stated in the discussion after (26)-(27), our goal in designing a separation method for complex noncircular sources

is to create an update whose analytical form follows that of (26). The first term in (41) is quite similar in form to (26), implying that the desired coefficient update before normalization should be defined as

$$\begin{aligned}\widetilde{\mathbf{c}}_t &= \mathbf{K}\,\mathrm{diag}\,\{\mathbf{c}_t\mathbf{c}_t^H\}\mathbf{c}_t \\ &= \widehat{\mathbf{M}}_t^T\mathbf{c}_t - \mathbf{c}_t[2\mathbf{c}^H\mathbf{c}] - \mathbf{\Lambda}\mathbf{c}_t^*[\mathbf{c}_t^T\mathbf{\Lambda}\mathbf{c}_t],\end{aligned} \tag{42}$$

where $\widehat{\mathbf{M}}_t$ is the expression in (40) with $\mathbf{c}_t$ replacing $\mathbf{c}$. Expressing this update in $\mathbf{w}_t$ coordinates gives

$$\widetilde{\mathbf{w}}_t = \mathbf{\Gamma}^*\widehat{\mathbf{M}}_t^T\mathbf{\Gamma}^T\mathbf{w}_t - \mathbf{w}_t[2\mathbf{w}_t^H\mathbf{w}_t] - \mathbf{\Gamma}^*\mathbf{\Lambda}\mathbf{\Gamma}^H\mathbf{w}_t[\mathbf{w}_t^T\mathbf{\Gamma}\mathbf{\Lambda}\mathbf{\Gamma}^T\mathbf{w}_t]. \tag{43}$$

Finally, we notice that

$$\mathbf{\Gamma}\mathbf{\Lambda}\mathbf{\Gamma}^T \approx \mathbf{P}_{VV} = \widehat{\mathbf{\Lambda}}, \tag{44}$$

$$\begin{aligned}\mathbf{\Gamma}^*\widehat{\mathbf{M}}_t^T\mathbf{\Gamma}^T\mathbf{w}_t &= \frac{1}{N}\sum_{n=1}^{N}\mathbf{v}^*(n)\mathbf{v}^T(n)\mathbf{w}_t\mathbf{w}_t^H\mathbf{v}^*(n)\mathbf{v}^T(n)\mathbf{w}_t \\ &= \frac{1}{N}\sum_{n=1}^{N}|y(n)|^2y(n)\mathbf{v}^*(n).\end{aligned} \tag{45}$$

Combining the above results gives the single-unit coefficient updates as

$$\widetilde{\mathbf{w}}_t = \left(\frac{1}{N}\sum_{n=1}^{N}|y(n)|^2y(n)\mathbf{v}^*(n)\right) - 2\mathbf{w}_t - \widehat{\mathbf{\Lambda}}\mathbf{w}_t^*[\mathbf{w}_t^T\widehat{\mathbf{\Lambda}}\mathbf{w}_t], \tag{46}$$

$$\mathbf{w}_{t+1} = \frac{\widetilde{\mathbf{w}}_t}{\sqrt{\widetilde{\mathbf{w}}_t^H\widetilde{\mathbf{w}}_t}}. \tag{47}$$

Remark 1. The above algorithm is similar in form to the FastICA algorithm for circular complex-valued sources in [15] for the choice $G(y) = (1/2)y^2$. The last term on the right-hand side of (46), however, is novel, and it is critical to obtaining good performance of the algorithm for noncircularly symmetric sources. Simulations in the next-to-last section verify this claim.

4.3. Algorithm based on ordinary prewhitening

The above algorithm requires the strong-uncorrelating transform for its implementation. Computing the strong-uncorrelating transform involves the Takagi factorization of a symmetric complex matrix. When the circularity coefficients $\{\lambda_i\}$ of $\mathbf{P}_{VV}$ are distinct, this factorization can be computed using the singular-value decomposition. The computation of the Takagi factorization in more-general scenarios, however, requires specialized numerical code. If the code for this factorization is not available, we offer an alternative implementation of our fixed-point algorithm for separating complex-valued noncircular sources which employs ordinary prewhitening. In this version, find any prewhitening matrix $\widehat{\mathbf{G}}$ such that

$$\widehat{\mathbf{G}}\mathbf{R}_{XX}\widehat{\mathbf{G}}^H = \mathbf{I}, \tag{48}$$

and set

$$\mathbf{v}(k) = \widehat{\mathbf{G}}(k)\mathbf{x}(k), \tag{49}$$

where

$$\widehat{\mathbf{P}} = \frac{1}{N}\sum_{n=1}^{N}\mathbf{v}(n)\mathbf{v}^T(n) = \widehat{\mathbf{G}}\mathbf{P}_{XX}\widehat{\mathbf{G}}^T. \tag{50}$$

Note that $\widehat{\mathbf{P}}$ will not be diagonal in general.

It is possible to retrace the steps taken to derive the updates in (46)-(47) under the assumption that $\widehat{\mathbf{P}}$ is not diagonal. These steps are straightforward and are omitted. The final version of the algorithm is

$$\widetilde{\mathbf{w}}_t = \left(\frac{1}{N}\sum_{n=1}^{N}|y(n)|^2y(n)\mathbf{v}^*(n)\right) - 2\mathbf{w}_t - \widehat{\mathbf{P}}^*\mathbf{w}_t^*[\mathbf{w}_t^T\widehat{\mathbf{P}}\mathbf{w}_t], \tag{51}$$

$$\mathbf{w}_{t+1} = \frac{\widetilde{\mathbf{w}}_t}{\sqrt{\widetilde{\mathbf{w}}_t^H\widetilde{\mathbf{w}}_t}}. \tag{52}$$

Remark 2. Comparing the updates in (46) and (51), we see that the price paid for not computing the Takagi factorization is an additional matrix-vector multiply within every iteration of the coefficient vector update. This computational increase is small relative to that needed to calculate $\mathbf{y}(n)$, $1 \le n \le N$, and the first term on the right-hand sides of (46) and (51), however, as these data-dependent terms make up the bulk of the computational requirements of the procedure.

4.4. Convergence of the single-unit algorithms

The overall goal in our design of fixed-point algorithms for separating complex-valued noncircular sources was to obtain procedures that exhibit the fast, globally convergent performance reminiscent of the algorithm in the real-valued case. Do the single-unit approaches in (46)-(47) and (51)-(52) achieve this end? The following theorem indicates that the answer is in the affirmative, the proof of which is in Appendix E.

Theorem 6. *As $N \to \infty$, both of the single-unit updates in* (46)-(47) *and* (51)-(52) *can be described in the combined system coefficient vector space as $\mathbf{c}_t = \mathbf{\Theta}_t\mathbf{a}_t$, where $\mathbf{\Theta}_t$ is a diagonal matrix of complex factors $\{e^{j\theta_i}[\mathrm{sgn}(\kappa_i)]^t\}$, $\mathbf{a}_t$ is a positive-valued m-dimensional vector obeying the relationships*

$$\begin{aligned}\widetilde{\mathbf{a}}_t &= \mathbf{K}_a\mathbf{F}(\mathbf{a}_t), \\ \mathbf{a}_{t+1} &= \frac{\widetilde{\mathbf{a}}_t}{\sqrt{\widetilde{\mathbf{a}}_t^T\widetilde{\mathbf{a}}_t}},\end{aligned} \tag{53}$$

where $\mathbf{K}_a$ is a diagonal matrix of the absolute values of the complex source kurtoses $\{|\kappa_1|, \ldots, |\kappa_m|\}$ with $\kappa_i = E\{|s_i(k)|^4\} - 2 - \lambda_i^2$, $\mathbf{F}(\mathbf{a}_t)$ is a diagonal matrix whose ith diagonal entry is a_{it}^3, and $\theta_i = \angle c_i(0)$. Thus, the convergence performance of either algorithms is mathematically identical to that of the real-valued FastICA algorithm with kurtosis contrast, where real-valued complex-source kurtoses replace real-source kurtoses and coefficient amplitudes replace the coefficient values in the evolutionary behavior.

Remark 3. The above result indicates that both of our single-unit algorithms do not attempt to change the phase of the separating solution during their operation, except for a trivial sign flip during odd-valued iterations. This attribute is highly desirable for practical applications, as it implies that separate procedures could be employed to extract the real and imaginary components of the sources in $\mathbf{s}(k)$ if $s_{R,i}(k)$ and $s_{I,i}(k)$ are statistically independent. This "phase-blind" behavior is obtained despite the fact that the underlying sources are potentially noncircular. Moreover, the algorithms also inherit the nice convergence properties of the FastICA algorithm in the real-valued mixture case [24–26].

5. FIXED-POINT ALGORITHMS FOR SEPARATING COMPLEX NONCIRCULAR SOURCE MIXTURES

To extend either of our proposed algorithms to general m-source extraction, we use similar concepts as in the real-valued FastICA algorithm extended to the complex realm. In particular, since $\mathbf{v}(k)$ is related to $\mathbf{s}(k)$ through the Hermitian matrix $\mathbf{\Gamma}$, then all m sources can be extracted by applying m versions of either algorithms to the sequence $\mathbf{v}(k)$ and constraining the resulting coefficient vectors to be complex orthogonal. This orthogonality could be maintained in one of two general recommended ways:[3]

(i) sequentially through a Gram-Schmidt or QR procedure, or

(ii) jointly through a symmetric orthogonalization procedure using an inverse matrix square root or an adaptive constraint method.

Sequential orthogonalization procedures that result in signal deflation are generally more robust to poor estimation of the contrast function and are provably convergent given enough measurements, but they suffer from error accumulation in the separation solutions such that sources extracted later in the procedure contain greater amounts of error and noise. Symmetric orthogonalization procedures provide higher separation performance when the sources can be well-identified via their non-Gaussian statistics but do not perform as well in other scenarios and are not guaranteed to converge for $m > 2$. To achieve the overall best performance, it is suggested that one designs algorithms that alternate between sequential and symmetric orthogonalization procedures to obtain both robust and accurate separation.

Algorithm 1 gives a sequential implementation of m versions of our proposed fixed-point algorithm for complex sources in (51)-(52), termed CFPA1, with Gram-Schmidt orthogonalization using the MATLAB technical computing environment. Algorithm 2 provides a parallel implementation of m versions of our proposed fixed-point algorithm for complex sources in (51)-(52), termed CFPA2, in which symmetric orthogonalization is used. Versions of the algorithm employing the updates in (46)-(47) and the strong-uncorrelated transform for prewhitening have been omitted but are simple to construct given the software for the Takagi factorization.

6. SIMULATIONS

We now explore the behaviors of our two fixed-point algorithms via Monte Carlo simulations. All of our evaluations are performed on synthetic data generated in the MATLAB technical computing environment to allow a straightforward evaluation and performance comparison between differing methods. In each case, we have used the average interchannel interference (ICI) to measure separation performance, which for the combined system matrix $\mathbf{C}_t = \mathbf{W}_t\hat{\mathbf{G}}\mathbf{A}$ with (i, j)th element c_{ijt} is given by

$$\mathrm{ICI}_t = \frac{1}{m}\sum_{i=1}^{m}\left(\frac{\sum_{l=1}^{m}|c_{ilt}|^2 - \max_{1\le k\le m}|c_{ikt}|^2}{\max_{1\le k\le m}|c_{ikt}|^2}\right). \quad (54)$$

This performance measure does not attempt to determine whether all sources are extracted individually, although the algorithms being compared enforce strong second-order orthogonality between the extracted outputs, making such an occurance extremely unlikely. An alternative to (54) is the Amari index [27]. The mixing matrix $\mathbf{A}$ has been generated randomly for each simulation run using an SVD-like combination of two random Hermitian matrices and a set of complex diagonal elements whose amplitudes were restricted to the interval $[0.2, 1]$. The random Hermitian matrices were generated by orthogonalizing the columns of square matrices with uncorrelated complex circular Gaussian elements. Both noiseless and noisy mixtures have been used, in which additive circular uncorrelated Gaussian noises with variances $\sigma_\nu^2 = 0.1$ were used as the measurement interference.

We compare the separation performance of our CFPA1 and CFPA2 algorithms to two different versions of two well-known existing methods for complex ICA: JADE [5], and the complex FastICA algorithm in [15] that assumes circularly symmetric source distributions, where an amplitude cost $G(|y|^2) = 0.5|y|^2$ has been used. All of the algorithms are simple to set up and require little effort in terms of parameter tuning. Even so, we employed two versions of JADE that involve simultaneous diagonalization of m and m^2 cumulant matrices, tuning the stopping parameters to obtain the best performance from each, as well as two versions of the FastICA algorithm in [15] employing symmetric orthogonalization and asymmetric deflation procedures, respectively.

[3] A third class of methods—adaptive orthogonalization through linear signal cancellation—is not recommended as it is generally not numerically robust.

```
function [B,y] = cfpa1(x);

[N,m] = size(x);
Rxx = (x'*x)/N;
[Q,Lam] = eig(Rxx);
Ghat = Q*diag(real(diag(Lam)).^ (-1/2));
v = x*Ghat;
Phat = (transpose(v)*v)/N;
W = eye(m);  y = zeros(N,m);
for i = 1:m
    k = 0;  Wold = zeros(m,1);
    Wt = W(:,i);
    while (abs(abs(Wold'*Wt)-1)>1e-4)*(k<100)
        k = k+1;
        Wold = Wt;
        yt = v*Wt;
        PhatW = Phat*Wt;
        Wt = (v'*(yt.*abs(yt). ^ 2))/N - 2*Wt - conj(PhatW)*(transpose(Wt)*PhatW);
        for n = 1:i-1
            Wt = Wt - W(:,n)*(W(:,n)'*Wt);
        end
        Wt = Wt/sqrt(Wt'*Wt);
    end
    y(:,i) = v*Wt;
    W(:,i) = Wt;
end
B = Ghat*W;
```

Algorithm 1: An implementation of our proposed fixed-point algorithm for complex-valued non-Gaussian source mixtures which uses sequential orthogonalization.

Since all of the algorithms being compared leverage the use of fourth-order source statistics, our study attempts to illuminate the advantages and weaknesses of the optimization methods used in each approach under finite-sample effects. One thousand evaluations of each method have been used to determine the averaged performance statistics shown.

Consider noiseless six-source mixtures of two real-valued binary-$\{\pm 1\}$ distributed sources, two 4QAM sources, and two 16QAM sources. Figure 1 shows the average ICI of the six algorithms tested as a function of data-block length N. As can be seen, our proposed methods perform better than either version of JADE and either version of the algorithm in [15] for small sample sizes, a result that is consistent throughout all of the results shown. The finite-sample performances of our proposed methods are quite good, offering separation of between 12.5 and 15 dB for only a block of 75 snapshots in this case. Because the mixture contains some real-valued sources, the complex FastICA procedure in [15] produces a biased result and is not competitive. The performances of the two JADE algorithms, and JADE(m^2) in particular, approach and exceed that of CFPA1 with asymmetric deflation, but CFPA2 with symmetric orthogonalization performs the best for all block lengths considered. As for repeatability, we evaluated the 95% confidence intervals for all six algorithms for all data points measured and we expressed the minimum and maximum of the ranges as ratios $r_{\min}$ and $r_{\max}$ of the average ICI in each case. The observed performance indicates that these confidence interval ratios do not change very much for different values of N, and Table 1 lists $E\{r_{\min}\}$ and $E\{r_{\max}\}$ for each algorithm. As can be seen, the repeatability of the proposed algorithms is similar to JADE(m) in this situation.

Additional experiments with both noiseless and noisy mixtures indicate that

(a) when a circularly symmetric complex Gaussian source is present, the roles of Algorithms 1 and 2 reverse, with the symmetric-orthogonalization-based CFPA2 technique performing the best;
(b) the proposed algorithms are robust to small amounts of low-level uncorrelated Gaussian observation noise (e.g., noise variances of $\sigma_n^2 = 0.001$ in the six-source scenario already considered).

We now consider a different source mixture scenario, in which we have used three source types—uniform-$[-\sqrt{3}, \sqrt{3}]$, unit-variance Laplacian, and binary—to generate nine different sources by (a) taking all possible pairs of the

```
function [B,y] = cfpa2(x);

[N,m] = size(x);
Rxx = (x'*x)/N;
[Q,Lam] = eig(Rxx);
Ghat = Q*diag(real(diag(Lam)).^ (-1/2));
v = x*Ghat;
Phat = (transpose(v)*v)/N;
W = eye(m); D = W; y = zeros(N,m); Wold = zeros(m); k = 0;
while (norm(abs(Wold'*W)-eye(m),'fro') >(m*1e-4))*(k<15*m)
    k = k+1;
    Wold = W;
    y = v*W;
    PhatW = Phat*W;
    for n = 1:m
        D(n,n) = transpose(W(:,n))*PhatW(:,n);
    end
    W = (v'*(y.* abs(y). ^ 2))/N - 2*W - conj(PhatW)*D;
    [Q,Lam] = eig(W'*W);
    W = W*(Q*diag(diag(real(Lam)). ^ (-1/2))*Q');
end
y = v*W;
B = Ghat*W;
```

Algorithm 2: An implementation of our proposed fixed-point algorithm for complex-valued non-Gaussian source mixtures which uses symmetric orthogonalization.

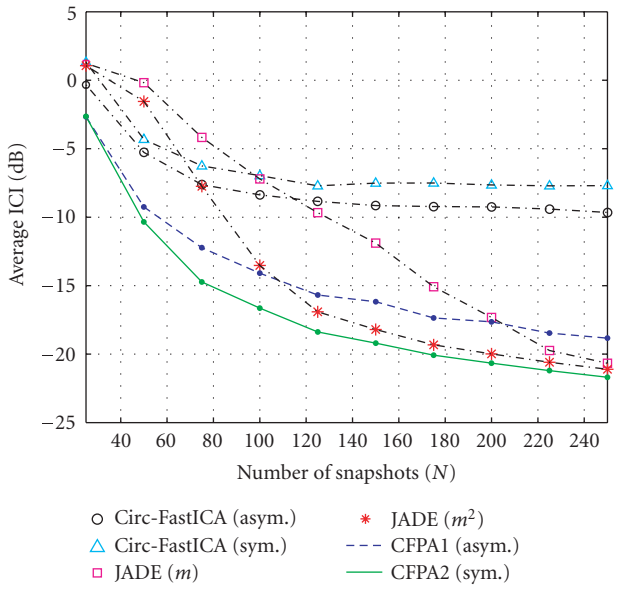

Figure 1: Average ICI as a function of data-record length N for the various algorithms on a noiseless six-source demixing task.

three real-valued distributions to create the real and imaginary parts of six complex sources, (b) including each of the three distributions as an additional real-valued source in the mixture, and (c) including a circularly symmetric Gaussian signal as part of the source signal set. Figure 2 shows the behaviors of the algorithms in this situation. The proposed methods are superior to existing ones for block sizes smaller than $N = 600$, and both of the proposed methods perform slightly better than JADE(m) for all block lengths considered. For larger block lengths, JADE(m^2) performs the best in this scenario.

The final source mixture scenario has complex-valued mixtures of six independent, identically distributed real-valued four-level (2B1Q) sources, in which uncorrelated zero-mean complex-valued jointly Gaussian observation noise with variance $\sigma_v^2 = 0.1$ has been added to each of the measurements. Due to the varying nature of the singular values of $\mathbf{A}$ within the measurements, the signal-to-noise ratios (SNRs) of the mixtures are simulation-run-dependent, but the minimum and maximum SNRs across all simulation runs are -4 dB and 10 dB, respectively, with an average SNR of 4 dB. Figure 3 shows the behaviors of the algorithms in this situation. Both of the proposed methods perform better than JADE(m) when fewer than 300 snapshots are available, and the performance of the CFPA1 method is only exceeded by that of JADE(m^2) for situations where more than 250 snapshots are available in this case.

In cases where the performance of our proposed methods are competitive with a joint-diagonalization approach

Table 1: Averaged 95% confidence intervals for the various algorithms as a ratio to the average ICI for the various algorithms in the first experiment.

Conf. interval ratio	JADE(m)	JADE(m^2)	Circ-FICA (asym.)	Circ-FICA(sym.)	CFPA (asym.)	CFPA (sym.)
$r_{\min}$	0.424	0.490	0.208	0.465	0.399	0.430
$r_{\max}$	2.58	1.98	2.20	1.87	2.44	2.11

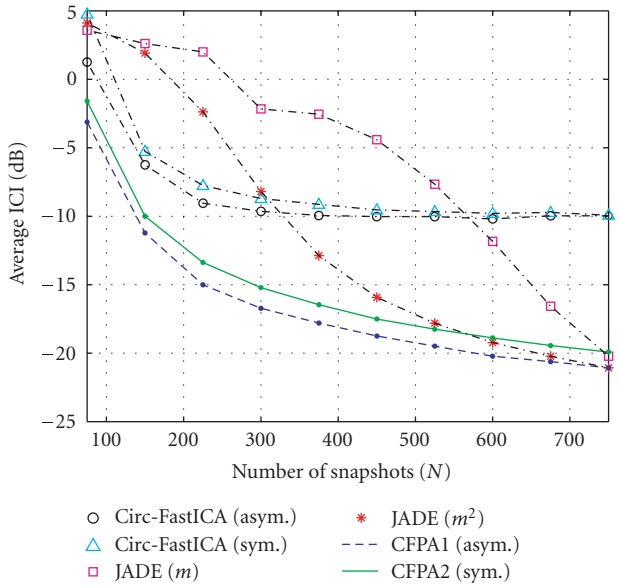

Figure 2: Average ICI as a function of data-record length N for the various algorithms on a more-challenging noiseless ten-source demixing task.

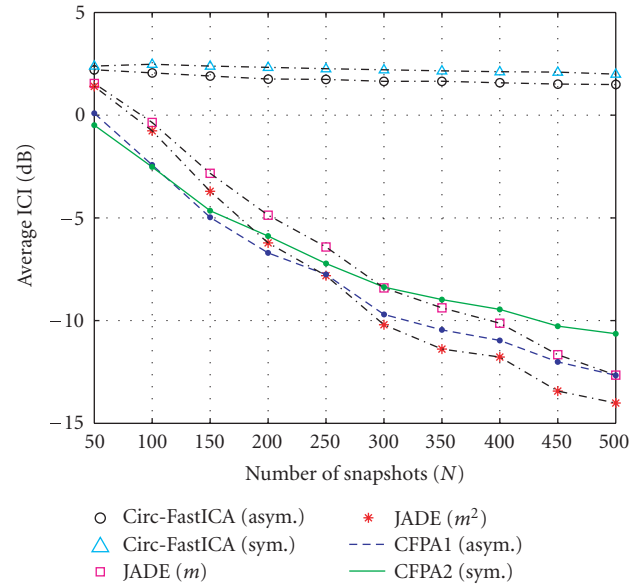

Figure 3: Average ICI as a function of data-record length N for the various algorithms on a noisy i.i.d. source separation task.

such as JADE, it is important to mention the computational advantages that the fixed-point approaches often provide. While both fixed-point algorithms and joint-diagonalization algorithms are iterative, it has been our observation that the fixed-point algorithms often complete their separation tasks more quickly than the joint-diagonalization algorithms when faced with large numbers of mixtures and/or large numbers of snapshots. In fact, it is both the slowness of the pairwise joint diagonalization procedure and the computational complexity of forming the cumulant estimates needed for JADE(m) and JADE(m^2) that prevented us from comparing the performance of these algorithms for large numbers of snapshots ($N \geq 10000$) and large numbers of channels ($m \geq 6$) on our computing equipment. On the other hand, we have successfully and repeatedly separated mixtures of $m = 25$ complex-valued sources with both the CFPA1 and CFPA2 algorithms using only a few seconds of CPU processing power on current-day PCs. The programs for these fixed-point methods generally run faster on modern computer hardware as well due to their use of sums-of-products calculations that are well supported in digital processors. Of course, it is possible to build specialized hardware to perform Givens rotations, so a system designer should select the algorithmic approach that makes the most sense for her or his preferred computational platform.

7. CONCLUSIONS

In this paper, we have carefully considered the design of blind source separation algorithms for mixtures of independent, noncircularly symmetric, and non-Gaussian sources. Using the structure of the symmetric fourth-order moment tensor of the source signal vector under strong-uncorrelation, we have constructed ICA algorithms that inherit all of the nice properties of the well-known kurtosis-contrast-based FastICA algorithm while being applicable to complex-valued signals. The techniques are computationally simple and employ well-known and well-understood data transformations such as whitening. Simulations indicate that the proposed techniques have finite-sample separation performance that usually meets or exceeds that of existing approaches for complex-valued blind source separation, especially for small data-record lengths. Extensions of these algorithmic methods to more-general and varied separation contrasts is the subject of current work.

APPENDICES

A. PROOF OF THEOREM 1

From the definition of $y(k)$, we have

$$\begin{aligned} y_R(k) &= c_R s_R(k) - c_I s_I(k), \\ y_I(k) &= c_I s_R(k) + c_R s_I(k). \end{aligned} \tag{A.1}$$

Define the quantities

$$\begin{aligned} \sigma_R^2 &= E\{s_R^2(k)\}, \\ \sigma_I^2 &= E\{s_I^2(k)\}, \\ \rho &= E\{s_R(k)s_I(k)\}. \end{aligned} \tag{A.2}$$

Then, by substitution of (A.1) into (5)-(6) and considering the real and imaginary parts of (6) separately, we have the three relations

$$(c_R^2 + c_I^2)(\sigma_R^2 + \sigma_I^2) = 1, \tag{A.3}$$

$$(c_R^2 - c_I^2)(\sigma_R^2 - \sigma_I^2) - 4c_R c_I \rho = \lambda, \tag{A.4}$$

$$(c_R^2 - c_I^2)\rho + c_R c_I(\sigma_R^2 - \sigma_I^2) = 0. \tag{A.5}$$

Let $c_R = A\cos\theta$ and $c_I = A\sin\theta$, where $A > 0$ and $0 \le \theta < \pi$. Then, the relation in (A.3) yields a solution for A as

$$A = \frac{1}{\sqrt{\sigma_R^2 + \sigma_I^2}}. \tag{A.6}$$

Considering the relations in (A.4) and (A.5), we can express them in terms of θ as

$$(\cos^2\theta - \sin^2\theta)\rho + \cos\theta\sin\theta(\sigma_R^2 - \sigma_I^2) = 0, \tag{A.7}$$

$$(\cos^2\theta - \sin^2\theta)\frac{\sigma_R^2 - \sigma_I^2}{\sigma_R^2 + \sigma_I^2} - 4\cos\theta\sin\theta\frac{\rho}{\sigma_R^2 + \sigma_I^2} = \lambda. \tag{A.8}$$

We can consider several cases based on the absolute value of ρ and the relative values of σ_R^2 and σ_I^2, finding solutions for θ and λ in the above relationship. For brevity, we only consider two specific cases below; the remaining cases are easily handled.

Case 1 (most specific case). Assume $\rho = 0$ and $\sigma_R^2 = \sigma_I^2$. Then, (A.7)-(A.8) are satisfied for any value of θ, and $\lambda = 0$.

Case 2 (most general case). Assume $\rho \neq 0$ and $\sigma_R^2 \neq \sigma_I^2$. Then, we can write (A.7) as

$$\cos^2\theta[(1 - \tan^2\theta)\rho + \tan\theta(\sigma_R^2 - \sigma_I^2)] = 0. \tag{A.9}$$

For values of θ in the range $0 < \theta < \pi$ not including $\theta \in \{\pi/2\}$, the above equation has two solutions given by

$$\tan\theta = \frac{\sigma_R^2 - \sigma_I^2}{2\rho} \pm \sqrt{\left(\frac{\sigma_R^2 - \sigma_I^2}{2\rho}\right)^2 + 1}. \tag{A.10}$$

The two solutions come about because of the two roots, where one root results in $\tan\theta > 0$ and the other root results in $\tan\theta < 0$. Thus, it is always possible to find a solution to A and θ satisfying (A.3) and (A.5) for $c = A\cos\theta + jA\sin\theta$. Now, considering (A.4) or its equivalent expression (A.8), we can write this relation in matrix form as

$$\begin{bmatrix}\cos\theta & \sin\theta\end{bmatrix}\begin{bmatrix} \dfrac{\sigma_R^2 - \sigma_I^2}{\sigma_R^2 + \sigma_I^2} & -\dfrac{2\rho}{\sigma_R^2 + \sigma_I^2} \\ -\dfrac{2\rho}{\sigma_R^2 + \sigma_I^2} & -\dfrac{\sigma_R^2 - \sigma_I^2}{\sigma_R^2 + \sigma_I^2} \end{bmatrix}\begin{bmatrix}\cos\theta \\ \sin\theta\end{bmatrix} = \lambda. \tag{A.11}$$

The (2×2) matrix in the above relation is symmetric with real-valued eigenvalues given by

$$r_{1,2} = \pm\sqrt{1 - 4\frac{\sigma_R^2\sigma_I^2 - \rho^2}{(\sigma_R^2 + \sigma_I^2)^2}}. \tag{A.12}$$

Because $\rho^2 \le \sigma_R^2\sigma_I^2$ by the Cauchy-Schwartz inequality, we have $|r_{1,2}| \le 1$, and because of the quadratic form of (A.11), this guarantees that

$$|\lambda| \le 1. \tag{A.13}$$

Finally, by combining (A.7) and (A.8), we can develop a second relation for λ in terms of θ as

$$[(\sigma_R^2 - \sigma_I^2)^2 + 4\rho^2]\left(\frac{\cos^2\theta}{\sigma_R^2 + \sigma_I^2}\right)\left[-\frac{\tan\theta}{\rho}\right] = \lambda. \tag{A.14}$$

By choosing the solution for θ in (A.10) that causes

$$\operatorname{sgn}(\tan\theta) = -\operatorname{sgn}(\rho), \tag{A.15}$$

we can guarantee that $\lambda > 0$. Combining the above results proves the theorem.

Proof of Corollary 1. Proving the relationships is straightforward by noting the following relations:

$$\begin{aligned} E\{(y_R(k))^2\} &= \frac{E\{|y(k)|^2\} + E\{[y(k)]^2\}}{2}, \\ E\{(y_I(k))^2\} &= \frac{E\{|y(k)|^2\} - E\{[y(k)]^2\}}{2}, \\ E\{y_R(k)y_I(k)\} &= \frac{\Im m[E\{[y(k)]^2\}]}{2}. \end{aligned} \tag{A.16}$$

□

Proof of Corollary 2. The relationship is obvious when considering the results of Corollary 1. □

B. PROOF OF THEOREM 2

Under the assumption that the elements of $\mathbf{s}(k)$ are zero-mean, independent, and strong-uncorrelated, we have that

$$\begin{aligned} E\{s_i(k)\} &= 0, \\ E\{|s_i(k)|^2\} &= 1, \\ E\{s_i(k)s_j^*(k)\} &= \begin{cases} 1, & i = j, \\ 0, & i \neq j, \end{cases} \\ E\{s_i(k)s_j(k)\} &= \begin{cases} \lambda_i, & i = j, \\ 0, & i \neq j, \end{cases} \end{aligned} \tag{B.1}$$

for $1 \le \{i, j\} \le m$. Under the above conditions, the expectation $E\{s_i(k)s_j^*(k)s_l^*(k)s_n(k)\}$ is nonzero only under the following four conditions.

(i) For $i = j \ne l = n$, $E\{s_i(k)s_j^*(k)s_l^*(k)s_n(k)\} = E\{|s_i(k)|^2\}E\{|s_l(k)|^2\} = 1$.
(ii) For $i = l \ne j = n$, $E\{s_i(k)s_j^*(k)s_l^*(k)s_n(k)\} = E\{|s_i(k)|^2\}E\{|s_j(k)|^2\} = 1$.
(iii) For $i = n \ne j = l$, $E\{s_i(k)s_j^*(k)s_l^*(k)s_n(k)\} = E\{s_i^2(k)\}(E\{s_j^2(k)\})^* = \lambda_i\lambda_j$.
(iv) For $i = j = k = n$, $E\{s_i(k)s_j^*(k)s_l^*(k)s_n(k)\} = E\{|s_i(k)|^4\}$.

Under any other index subset of i, j, k, and l,

$$E\{s_i(k)s_j^*(k)s_l^*(k)s_n(k)\} = 0. \tag{B.2}$$

In particular, situations in which $i = j = k \ne n$, $i = j = n \ne k$, $i = k = n \ne j$, and $j = k = n \ne i$ result in a zero moment because of the zero means and statistical independence of the elements of $\mathbf{s}(k)$. The theorem then follows by using the definition of κ_i in (11).

Proof of Corollary 3. Consider the value of $E\{|s_i(k)|^4\}$ under the assumption that $s_i(k)$ is strong-uncorrelated Gaussian-distributed. Then,

$$\begin{aligned}
E\{|s_i(k)|^4\} &= \iint_{-\infty}^{\infty} (s_R^2 + s_I^2)^2 \frac{1}{\pi\sqrt{1-\lambda_i^2}} \\
&\qquad \times \exp\left(-\left[\frac{s_R^2}{(1+\lambda_i)} + \frac{s_I^2}{(1-\lambda_i)}\right]\right) ds_R ds_I \\
&= \int_{-\infty}^{\infty} \frac{s_R^4}{\sqrt{2\pi((1+\lambda_i)/2)}} \exp\left(-\frac{s_R^2}{(1+\lambda_i)}\right) ds_R \\
&\quad \times \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi((1-\lambda_i)/2)}} \exp\left(-\frac{s_I^2}{(1-\lambda_i)}\right) ds_I \\
&\quad + 2\int_{-\infty}^{\infty} \frac{s_R^2}{\sqrt{2\pi((1+\lambda_i)/2)}} \exp\left(-\frac{s_R^2}{(1+\lambda_i)}\right) ds_R \\
&\quad \times \int_{-\infty}^{\infty} \frac{s_I^2}{\sqrt{2\pi((1-\lambda_i)/2)}} \exp\left(-\frac{s_I^2}{(1-\lambda_i)}\right) ds_I \\
&\quad + \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi((1+\lambda_i)/2)}} \exp\left(-\frac{s_R^2}{(1+\lambda_i)}\right) ds_R \\
&\quad \times \int_{-\infty}^{\infty} \frac{s_I^4}{\sqrt{2\pi((1-\lambda_i)/2)}} \exp\left(-\frac{s_I^2}{(1-\lambda_i)}\right) ds_I.
\end{aligned} \tag{B.3, B.4}$$

Consider first the case that $0 \le \lambda < 1$. Then, by the change of variables

$$\begin{aligned}
u_R &= \sqrt{\frac{2}{(1+\lambda_i)}}\, s_R, \\
u_I &= \sqrt{\frac{2}{(1-\lambda_i)}}\, s_I,
\end{aligned} \tag{B.5}$$

we can rewrite (B.4) as

$$\begin{aligned}
E\{|s_i(k)|^4\} &= \left(\frac{1+\lambda_i}{2}\right)^2 \int_{-\infty}^{\infty} \frac{u_R^4}{\sqrt{2\pi}} \exp\left(-\frac{u_R^2}{2}\right) du_R \\
&\quad \times \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u_I^2}{2}\right) du_I \\
&\quad + \frac{1-\lambda_i^2}{2} \int_{-\infty}^{\infty} \frac{u_R^2}{\sqrt{2\pi}} \exp\left(-\frac{u_R^2}{2}\right) du_R \\
&\quad \times \int_{-\infty}^{\infty} \frac{u_I^2}{\sqrt{2\pi}} \exp\left(-\frac{u_I^2}{2}\right) du_I \\
&\quad + \left(\frac{1-\lambda_i}{2}\right)^2 \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u_R^2}{2}\right) du_R \\
&\quad \times \int_{-\infty}^{\infty} \frac{u_I^4}{\sqrt{2\pi}} \exp\left(-\frac{u_I^2}{2}\right) du_I.
\end{aligned} \tag{B.6}$$

From the properties of Gaussian random variables,

$$\begin{aligned}
\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right) du &= 1, \\
\int_{-\infty}^{\infty} \frac{u^2}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right) du &= 1, \\
\int_{-\infty}^{\infty} \frac{u^4}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right) du &= 3,
\end{aligned} \tag{B.7}$$

such that (B.6) becomes

$$\begin{aligned}
E\{|s_i(k)|^4\} &= 3\left[\left(\frac{1+\lambda_i}{2}\right)^2 + \left(\frac{1-\lambda_i}{2}\right)^2\right] + \frac{1-\lambda_i^2}{2} \\
&= \frac{1}{4}[3(1+2\lambda_i+\lambda_i^2)+3(1-2\lambda_i+\lambda_i^2)+2(1-\lambda_i^2)] \\
&= 2 + \lambda_i^2.
\end{aligned} \tag{B.8}$$

Consider now the limiting case that $\lambda \to 1$. Then it can be shown that the function

$$f_\lambda(s) = \frac{1}{\sqrt{2\pi((1-\lambda_i)/2)}} \exp\left(-\frac{s_I^2}{(1-\lambda_i)}\right) \tag{B.9}$$

is a family of generalized functions with unit area [28], such that

$$\begin{aligned}
\lim_{\lambda\to 1} f_\lambda(s) &= \delta(s), \\
\lim_{\lambda\to 1} \int_{-\infty}^{\infty} s^p f_\lambda(s) ds &= 1
\end{aligned} \tag{B.10}$$

for any nonnegative integer p. Thus, when $\lambda = 1$, (B.6) reduces to

$$E\{|s_i(k)|^4\} = \int_{-\infty}^{\infty} \frac{s_R^4}{\sqrt{2\pi}} \exp\left(-\frac{s_R^2}{2}\right) ds_R = 3, \quad \text{(B.11)}$$

where $2 + \lambda_i = 3$ in this case. Considering $0 \leq \lambda \leq 1$, the result follows. □

C. PROOF OF THEOREM 3

The proof relies on the following expectations:

$$\begin{aligned} E\{\mathbf{s}(k)\} &= \mathbf{0}, \\ E\{\mathbf{s}(k)\mathbf{s}^H(k)\} &= \mathbf{I}, \\ E\{\mathbf{s}(k)\mathbf{s}^T(k)\} &= \mathbf{\Lambda}, \end{aligned} \quad \text{(C.1)}$$

as well as the symmetric fourth-moment tensor expression in (9). Since $y(k) = \mathbf{c}^T\mathbf{s}(k)$, we can easily see that

$$\begin{aligned} E\{y(k)\} &= \mathbf{c}^T E\{\mathbf{s}(k)\} = \mathbf{0}, \\ E\{|y(k)|^2\} &= \mathbf{c}^T E\{\mathbf{s}(k)\mathbf{s}^H(k)\}\mathbf{c}^* = \mathbf{c}^T\mathbf{c}^* = \sum_{i=1}^{m} |c_i|^2, \\ E\{y^2(k)\} &= \mathbf{c}^T E\{\mathbf{s}(k)\mathbf{s}^T(k)\}\mathbf{c} = \mathbf{c}^T\mathbf{\Lambda}\mathbf{c} = \sum_{i=1}^{m} \lambda_i c_i^2. \end{aligned} \quad \text{(C.2)}$$

Considering now the fourth-moment term in (20), we can write this in matrix form as

$$E\{|y(k)|^4\} = \mathbf{c}^T\mathbf{M}\mathbf{c}^*, \quad \text{(C.3)}$$

where the matrix $\mathbf{M}$ is

$$\mathbf{M} = E\{\mathbf{s}(k)\mathbf{s}^H(k)\mathbf{c}^*\mathbf{c}^T\mathbf{s}(k)\mathbf{s}^H(k)\}. \quad \text{(C.4)}$$

The (i,l)th entry of this matrix is

$$[\mathbf{M}]_{il} = \sum_{p=1}^{m}\sum_{q=1}^{m} E\{s_i(k)s_l^*(k)s_p^*(k)s_q(k)\}c_p^*c_q. \quad \text{(C.5)}$$

Using the properties of the fourth-moment tensor in (9), we have

$$[\mathbf{M}]_{il} = \begin{cases} \displaystyle\sum_{p=1, p\neq i}^{m} |c_p|^2 + E\{|s_i(k)|^4\}|c_i|^2 & \text{if } i = l, \\ c_i^*c_l + \lambda_i\lambda_l c_i c_l^* & \text{if } i \neq j. \end{cases} \quad \text{(C.6)}$$

Expressing this answer in matrix form,

$$\mathbf{M} = \mathbf{c}^*\mathbf{c}^T + \mathbf{I}\mathbf{c}^H\mathbf{c} + \mathbf{\Lambda}\mathbf{c}\mathbf{c}^H\mathbf{\Lambda} + \mathbf{N}, \quad \text{(C.7)}$$

where $\mathbf{N}$ is a diagonal matrix whose ith diagonal entry is

$$[\mathbf{N}]_{ii} = |c_i|^2(E\{|s_i(k)|^4\} - 2 - \lambda_i^2). \quad \text{(C.8)}$$

Recognizing the form of the symmetric kurtosis in (11), we can evaluate the quadratic form $\mathbf{c}^T\mathbf{M}\mathbf{c}^*$ in (C.3), which yields the expression in (20).

Proof of Corollary 4. Recognizing from the definition of signal kurtosis that

$$\kappa[y(k)] = E\{|y(k)|^4\} - 2(E\{|y(k)|^2\})^2 - |E\{y_i^2(k)\}|^2, \quad \text{(C.9)}$$

the expression in (21) is easily obtained by substituting the moment relations of (18), (19), and (20) into (C.9). □

D. PROOF OF THEOREM 5

By the law of large numbers, as $N \to \infty$, the summation in (40) converges to the statistical expectation

$$\lim_{N\to\infty} \widehat{\mathbf{M}} = E\{\mathbf{s}(k)\mathbf{s}^H(k)\mathbf{c}^*\mathbf{c}^T\mathbf{s}(k)\mathbf{s}^H(k)\}. \quad \text{(D.1)}$$

The expression on the right-hand side is identical to that in (C.4), for which its evaluated value is given by (C.7). Finally, noting that $\mathbf{N} = \mathbf{K}\,\mathrm{diag}\{\mathbf{c}\mathbf{c}^H\}$, the theorem follows.

E. PROOF OF THEOREM 6

Considering both (46)-(47) and (51)-(52) jointly, we premultiply the left- and right-hand sides of all expressions by $\mathbf{\Gamma}^T$. Taking $N \to \infty$ and making use of the relationships in (36) and (44), we obtain the common evolutionary equations

$$\begin{aligned} \widetilde{\mathbf{c}}_t &= \mathbf{K}\,\mathrm{diag}\,[\mathbf{c}\mathbf{c}^H]\mathbf{c}, \\ \mathbf{c}_{t+1} &= \frac{\widetilde{\mathbf{c}}_t}{\sqrt{\widetilde{\mathbf{c}}_t^H\widetilde{\mathbf{c}}_t}}. \end{aligned} \quad \text{(E.1)}$$

Define the combined system coefficient vector using complex phasor representation as

$$c_{it} = A_{it}e^{j\theta_t}. \quad \text{(E.2)}$$

Then, the update relations in (E.1) can be written in scalar form as

$$A_{i(t+1)}e^{j\theta_{i(t+1)}} = \frac{|\kappa_i|A_{it}^3 e^{j\theta_{it}}\,\mathrm{sgn}\,[\kappa_i]}{\sqrt{\sum_{p=1}^{m}\kappa_i^2 A_{pt}^6}}. \quad \text{(E.3)}$$

Clearly, we have $\theta_{i(t+1)} = \theta_{it} + \pi = \theta_i(0) + \pi t$ corresponding to a simple sign change of the ith coefficient if $\kappa_i < 0$ and $\theta_{i(t+1)} = \theta_{it} = \theta_i(0)$ corresponding to no sign change of the ith coefficient if $\kappa_i > 0$. Finally, defining the real-valued vector $\mathbf{a}_t = [A_{1t} \;\cdots\; A_{mt}]^T$, the evolutionary behavior of $\mathbf{a}_t$ follows the equations in (53), which are identical in form to the evolutionary equations defining the behavior of the m-dimensional real-valued single-unit FastICA procedure with kurtosis contrast, as derived in [7, 24].

ACKNOWLEDGMENTS

The author would like to thank Jan Eriksson and Visa Koivunen for various helpful discussions during the course of writing this manuscript.

REFERENCES

[1] A. Cichocki, R. Unbehauen, and E. Rummert, "Robust learning algorithm for blind separation of signals," *Electronics Letters*, vol. 30, no. 17, pp. 1386–1387, 1994.

[2] A. J. Bell and T. J. Sejnowski, "An information-maximization approach to blind separation and blind deconvolution," *Neural Computation*, vol. 7, no. 6, pp. 1129–1159, 1995.

[3] S. Amari, A. Cichocki, and H. H. Yang, "A new learning algorithm for blind signal separation," in *Advances in Neural Information Processing Systems*, vol. 8, pp. 757–763, MIT Press, Cambridge, Mass, USA, 1996.

[4] D. T. Pham, "Blind separation of instantaneous mixture of sources via an independent component analysis," *IEEE Transactions on Signal Processing*, vol. 44, no. 11, pp. 2768–2779, 1996.

[5] J. F. Cardoso and A. Souloumiac, "Blind beamforming for non-Gaussian signals," *IEE Proceedings, Part F: Radar and Signal Processing*, vol. 140, no. 6, pp. 362–370, 1993.

[6] P. Comon, "Independent component analysis. A new concept?" *Signal Processing*, vol. 36, no. 3, pp. 287–314, 1994.

[7] A. Hyvärinen and E. Oja, "A fast fixed-point algorithm for independent component analysis," *Neural Computation*, vol. 9, no. 7, pp. 1483–1492, 1997.

[8] L. Molgedey and H. G. Schuster, "Separation of a mixture of independent signals using time delayed correlations," *Physical Review Letters*, vol. 72, no. 23, pp. 3634–3637, 1994.

[9] L. Tong, R.-W. Liu, V. C. Soon, and Y.-F. Huang, "Indeterminacy and identifiability of blind identification," *IEEE Transactions on Circuits and Systems*, vol. 38, no. 5, pp. 499–509, 1991.

[10] A. Belouchrani, K. Abed-Meraim, J.-F. Cardoso, and E. Moulines, "A blind source separation technique using second-order statistics," *IEEE Transactions on Signal Processing*, vol. 45, no. 2, pp. 434–444, 1997.

[11] S. Amari, S. C. Douglas, A. Cichocki, and H. H. Yang, "Multichannel blind deconvolution and equalization using the natural gradient," in *Proceedings of the 1st IEEE Signal Processing Workshop on Signal Processing Advances in Wireless Communications (SPAWC '97)*, pp. 101–104, Paris, France, April 1997.

[12] T. Ristaniemi and J. Joutsensalo, "Advanced ICA-based receivers for block fading DS-CDMA channels," *Signal Processing*, vol. 82, no. 3, pp. 417–431, 2002.

[13] V. Calhoun and T. Adali, "Complex infomax: convergence and approximation of infomax with complex nonlinearities," in *Proceedings of the 12th IEEE Workshop on Neural Networks for Signal Processing*, pp. 307–316, Martigny, Switzerland, September 2002.

[14] J. Anemüller, T. J. Sejnowski, and S. Makeig, "Complex independent component analysis of frequency-domain electroencephalographic data," *Neural Networks*, vol. 16, no. 9, pp. 1311–1323, 2003.

[15] E. Bingham and A. Hyvärinen, "A fast fixed-point algorithm for independent component analysis of complex valued signals," *International Journal of Neural Systems*, vol. 10, no. 1, pp. 1–8, 2000.

[16] L. De Lathauwer and B. De Moor, "On the blind separation of non-circular sources," in *Proceedings of the 11th European Signal Processing Conference (EUSIPCO '02)*, Toulouse, France, September 2002.

[17] J. Eriksson and V. Koivunen, "Complex-valued ICA using second order statistics," in *Proceedings of the 14th IEEE Signal Processing Society Workshop on Machine Learning for Signal Processing*, pp. 183–191, Sao Luis, Brazil, September-October 2004.

[18] J. Eriksson and V. Koivunen, "Complex random vectors and ICA models: identifiability, uniqueness, and separability," *IEEE Transactions on Information Theory*, vol. 52, no. 3, pp. 1017–1029, 2006.

[19] M. Novey and T. Adali, "ICA by maximization of nongaussianity using complex functions," in *Proceedings of IEEE Workshop on Machine Learning for Signal Processing*, pp. 21–26, Mystic, Conn, USA, September 2005.

[20] J. Eriksson, A.-M. Seppola, and V. Koivunen, "Complex ICA for circular and non-circular sources," in *Proceedings of the 13th European Signal Processing Conference (EUSIPCO '05)*, Antalya, Turkey, September 2005.

[21] O. Shalvi and E. Weinstein, "Super-exponential methods for blind deconvolution," *IEEE Transactions on Information Theory*, vol. 39, no. 2, pp. 504–519, 1993.

[22] S.-Y. Kung, "Independent component analysis in hybrid mixture: KuicNet learning algorithm and numerical analysis," in *Proceedings of International Symposium on Multimedia Information Processing*, pp. 368–381, Taipei, Taiwan, December 1997.

[23] P. A. Regalia and M. Mboup, "Undermodeled equalization: a characterization of stationary points for a family of blind criteria," *IEEE Transactions on Signal Processing*, vol. 47, no. 3, pp. 760–770, 1999.

[24] S. C. Douglas, "On the convergence behavior of the FastICA algorithm," in *Proceedings of the 4th International Symposium on Independent Component Analysis and Blind Signal Separation*, pp. 409–414, Kyoto, Japan, April 2003.

[25] S. C. Douglas, "A statistical convergence analysis of the FastICA algorithm for two-source mixtures," in *Proceedings of the 39th Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, Calif, USA, October 2005.

[26] S. C. Douglas, Z. Yuan, and E. Oja, "Average convergence behavior of the FastICA algorithm for blind source separation," in *Proceedings of the 6th International Conference on Independent Component Analysis and Blind Signal Separation (ICA '06)*, vol. 3889, pp. 790–798, Charleston, SC, USA, March 2006.

[27] A. Cichocki and S. Amari, *Adaptive Blind Signal and Image Processing: Learning Algorithms and Applications*, John Wiley & Sons, New York, NY, USA, 2002.

[28] R. Bracewell, *The Fourier Transform and Its Applications*, McGraw-Hill, New York, NY, USA, 3rd edition, 1999.

Scott C. Douglas is an Associate Professor in the Department of Electrical Engineering at Southern Methodist University, Dallas, TX, and the Associate Director for the Institute for Engineering Education at SMU. He received his B.S., M.S., and Ph.D. degrees from Stanford University. He is a recognized expert in the fields of adaptive filters, blind source separation, and active noise control, having authored or coauthored two books, six book chapters, and over 150 papers in journals and conference proceedings. He is a recipient of an NSF Career (Young Investigator) Award and has received significant research funding from the US Army, DARPA, other US governmental

organizations, the State of Texas, and numerous companies. He is highly active in professional societies and has served as an Associate Editor for both the IEEE Transactions on Signal Processing and the IEEE Signal Processing Letters. He has served on the organizing committees of numerous international conferences and workshops as Technical Chair, Publications Chair, and Exhibits Chair, and is the General Chair of the 2010 International Conference on Acoustics, Speech, and Signal Processing. He has given many keynote and invited lectures as well as short courses on topics ranging from adaptive signal processing and control to innovative engineering education methods. Most recently, he has coauthored textbooks and developed materials and technology for the Infinity Project, a multifaceted effort to establish a United States engineering curriculum at precollege educational levels. He is a frequent consultant to industry, a Senior Member of the IEEE, and a Member of both Phi Beta Kappa and Tau Beta Pi.