FIXED-POINT FASTICA ALGORITHMS FOR THE BLIND SEPARATION OF COMPLEX-VALUED SIGNAL MIXTURES

Scott C. Douglas

Department of Electrical Engineering
Southern Methodist University
Dallas, Texas 75275 USA

ABSTRACT

In this paper, we derive new fixed-point algorithms for the blind separation of complex-valued mixtures of independent, possibly non-circularly-symmetric, and non-Gaussian source signals. Leveraging recent results in complex independent component analysis, we construct iterative procedures for complex signal mixtures whose evolutionary characteristics are identical to those of the real-valued FastICA algorithm with kurtosis contrast. Our methods are fast, computationally simple, and easy to use. 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 often have equal or better finite-sample performance than existing separation methods for complex-valued mixtures.

1. INTRODUCTION

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

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

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, the sources $\{s_i(k)\}$ are assumed to be statistically-independent and real-valued, and $\mathbf{A}$ is full rank. If certain additional conditions are met, it is possible to estimate demixing matrix $\mathbf{B}$ from $\mathbf{x}(k)$ such that

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

contains demixed independent elements that are possibly scaled and shuffled with respect to the sources in $\mathbf{s}(k)$. For real-valued mixtures, numerous algorithms have been developed [1]–[3]. Among these methods, the FastICA procedure in [3] has fast convergence, global convergence for kurtosis-based contrasts, and no step size parameter.

Suppose that $\mathbf{A}$ 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 in wireless communications, array processing, and medicine [4]–[6]. Fewer algorithms for separating complex signal mixtures have been described in the scientific literature [1, 4, 5, 7, 8]. The super-exponential algorithm in [7] involves pseudo-inverses of covariance matrices that may be difficult to calculate. In [8], the complex-valued source signals are 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 structure that is distinct from the real-valued case [9, 10]. For example, the matrix $\mathbf{A}$ can sometimes be identified up to scaling and permutation factors if $\mathbf{s}(k)$ contains multiple complex non-circular *Gaussian*-distributed sources. The key concept in these results is the relaxing of the circularity assumptions of the distributions of the complex sources $\{s_i(k)\}$. Algorithms for separating general-form complex source mixtures have appeared only recently [10], 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 task for mixtures of non-Gaussian and possibly non-circular source signals. The role of decorrelation in complex-valued ICA is carefully delineated, taking account of the results in [10]. We then present extensions of the popular FastICA algorithm for fourth-moment separation criteria to the non-circular complex-valued case. Unlike [8], our derivation exploits the structure of the fourth-moment symmetric tensor of the source signal vector directly. When applied to mixtures of non-circular and/or circular complex-valued sources, our algorithms preserve the fast and efficient convergence properties of the original FastICA algorithm for real-valued signal mixtures [3]. Brief convergence proofs of the algorithms are outlined showing that they work when $\mathbf{s}(k)$ contains at least $(m-1)$ non-Gaussian-distributed sources. Simulations indicate their separating capabilities for complex-valued BSS tasks.

2. COMPLEX-VALUED RANDOM VARIABLES

In our derivation, we first delineate the statistical structure of complex-valued random variables. We shall later use this structure to develop efficient separation algorithms. Let $s(k) = s_R(k) + js_I(k)$ denote a scalar complex-valued ran-

1-4244-0132-1/05/$20.00 ©2005 IEEE

dom 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 $p_R(s_R)$ and $p_I(s_I)$. Let $E\{g(s(k))\}$ denote the expectation operator for an arbitrary complex function $g(s(k)) = g_R(s_R(k), s_I(k)) + jg_I(s_R(k), s_I(k))$. The complex conjugate of $s(k)$ is $s^*(k) = s_R(k) - js_I(k)$. For convenience, $s(k)$ is a zero-mean random variable, *i.e.* $E\{s(k)\} = E\{s_R(k)\} = E\{s_I(k)\} = 0$.

Let $y(k) = cs(k)$, where $c = c_R + jc_I$ is a complex scalar. Clearly, $E\{y(k)\} = 0$. The following theorem relate to the distribution of $y(k)$, the proof of which is in [15].

Theorem 1: *For any zero mean complex r.v. $s(k)$, it is always possible to find a complex scalar c such that*

$$E\{|y(k)|^2\} = 1 \quad (3)$$

$$E\{[y(k)]^2\} = \lambda, \quad (4)$$

where the circularity coefficient λ is a real number satisfying $0 \leq \lambda \leq 1$. Furthermore, $E\{y_R^2(k)\} \geq E\{y_I^2(k)\}$, with equality if and only if $y(k)$ is proper, i.e., $E\{[y(k)]^2\} = 0$.

The above theorem states 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 no 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 in [10]. Note that this structure says nothing about the dependence of $s_R(k)$ and $s_I(k)$ or the distribution of $s(k)$. We now assume that $s_i(k)$ possesses this structure, as we can absorb the scaling of each source into the mixing matrix $\mathbf{A}$ in (1).

Our algorithms exploit the fourth-order moment structure of the vector $\mathbf{s}(k)$. Fourth-order cumulants have been heavily-used in deriving blind identification and separation algorithms in the real-valued case. A theorem and corollary give the fourth-order moment properties of strong-uncorrelated independent sources $\{s_i(k)\}$. Proofs are in [15].

Theorem 2: *Let $\mathbf{s}(k)$ contain m zero-mean, independent, strong-uncorrelated signals $s_i(k)$ with circularity coefficients λ_i, $0 \leq \lambda_i \leq 1$. Then, the symmetric fourth-order moment tensor $K_{ilnp} = E\{s_i(k)s_l^*(k)s_n^*(k)s_p(k)\}$ and symmetric kurtosis $\kappa_i = E\{|s_i(k)|^4\} - 2(E\{|s_i(k)|^2\})^2 - \left|E\{s_i^2(k)\}\right|^2$ are*

$$K_{ilnp} = \begin{cases} 1 & \text{if } i = l \neq n = p \\ & \text{or } i = n \neq l = p \\ \lambda_i\lambda_l & \text{if } i = p \neq l = n \\ \kappa_i + 2 + \lambda_i^2 & \text{if } i = l = n = p \\ 0 & \text{otherwise} \end{cases}, \quad (5)$$

$$\kappa_i = E\{|s_i(k)|^4\} - 2 - \lambda_i^2. \quad (6)$$

Corollary 2.1: If $s_i(k)$ is a strong-uncorrelated Gaussian random variable, its symmetric kurtosis κ_i is zero.

Because of the importance of the kurtosis in our work, we shall define the kurtosis operator for $s(k)$ as

$$\kappa[s(k)] = E\{|s(k)|^4\} - 2(E\{|s_i(k)|^2\})^2 - \left|E\{s_i^2(k)\}\right|^2 \quad (7)$$

The symmetric fourth-order moment tensor K_{ilnp} 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 the third-order moments vanish in all three cases. For independent $\{s_i(k)\}$ in the non-circular complex case, however, only *strong-uncorrelated* random variables maintain this nice structure for the tensor K_{ilnp}.

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

Consider a single row $\mathbf{b}^T = [b_1 \cdots b_m]^T$ of the separation matrix $\mathbf{B}$ that extracts a single source $s_i(k)$. The output signal at time k is $y(k) = \mathbf{b}^T\mathbf{x}(k)$. We can write $y(k)$ in terms of the combined coefficient vector $\mathbf{c} = \mathbf{A}^T\mathbf{b}$ as

$$y(k) = \mathbf{c}^T\mathbf{s}(k). \quad (8)$$

The next theorem describes the moments of $y(k)$ [15]:

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

$$E\{|y(k)|^4\} = \sum_{i=1}^{m} \kappa_i|c_i|^4 + 2\left(\sum_{i=1}^{m}|c_i|^2\right)^2 + \left(\sum_{i=1}^{m}\lambda_i c_i^2\right) \quad (9)$$

$$E\{|y(k)|^2\} = \sum_{i=1}^{m}|c_i|^2, \quad E\{[y(k)]^2\} = \sum_{i=1}^{m}\lambda_i c_i^2. \quad (10)$$

Corollary 3.1: The kurtosis of $y(k)$ is

$$\kappa[y(k)] = \sum_{i=1}^{m}\kappa_i|c_i|^4. \quad (11)$$

The result in (11) indicates two important facts:

1. For strong-uncorrelated sources, the kurtosis of $y(k)$ in the combined coefficient space depends on the circularity coefficients $\{\lambda_i\}$ only through the values κ_i in (6).

2. Let $c_i = A_i e^{j\theta_i}$. Then, using this polar representation,

$$\kappa[y(k)] = \sum_{i=1}^{m}\kappa_i A_i^4 \quad \text{and} \quad E\{|y(k)|^2\} = \sum_{i=1}^{m}A_i^2. \quad (12)$$

The relations in (12) have been used to develop algorithms for single-channel blind deconvolution (*c.f.* [11]) and for blind source separation of real-valued signal mixtures (*c.f.* [2]). In the latter case, the real-valued combined system coefficients play roles that are identical to those of the amplitudes of the $\{c_i\}$ in the complex-valued case. It is this latter correspondence that allows us to directly state an optimization strategy for extracting a single source [15]:

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

$$\mathcal{J}(\mathbf{b}) = \left| \frac{\kappa[y(k)]}{(E\{|y(k)|^2\})^2} \right|, \qquad (13)$$

where $y(k) = \mathbf{b}^T\mathbf{x}(k)$*. Assume that at least one of the sources has a non-zero 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.*

4. SINGLE-UNIT FIXED-POINT ALGORITHMS FOR COMPLEX MIXTURES

In [3], 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 contrast is used, the algorithm has a particularly appealing form in the combined system coefficient vector $\mathbf{c}_t = \mathbf{A}^T\mathbf{b}_t$ at iteration t [3, 12]:

$$\widetilde{\mathbf{c}}_t = \mathbf{K}\mathbf{F}(\mathbf{c}_t) \qquad (14)$$

$$\mathbf{c}_{t+1} = \frac{\widetilde{\mathbf{c}}_t}{\sqrt{\widetilde{\mathbf{c}}_t^T\widetilde{\mathbf{c}}_t}}, \qquad (15)$$

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. The utility of this algorithm can be inferred from (14)–(15), which leads to cubic convergence near a separating solution. Moreover, its average limiting performance over a uniform prior of initial coefficient vector directions is exponential with a rate of $(1/3)$ [12, 13, 14]. In what follows, we develop algorithms whose coefficient updates obey a similar relation as (14)–(15) 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. 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). The elements of $\mathbf{s}(n)$ are realizations of m statistically-independent complex-valued possibly non-circular random processes, in which only one source has a zero kurtosis.

Our derivation relies on the strong-uncorrelating transform for signal prewhitening [9]. The strong-uncorrelating transform is a linear transformation that diagonalizes both the covariance matrix and pseudo-covariance matrix

$$\mathbf{R}_{XX} = \frac{1}{N}\sum_{n=1}^{N}\mathbf{x}(n)\mathbf{x}^H(n) \qquad (16)$$

$$\mathbf{P}_{XX} = \frac{1}{N}\sum_{n=1}^{N}\mathbf{x}(n)\mathbf{x}^T(n), \qquad (17)$$

respectively, where $\cdot^H$ denotes Hermitian (conjugate) transpose. Note that $\mathbf{P}_{XX} \neq \mathbf{0}$ for finite sample sizes, and it remains non-zero as $N \to \infty$ if any of the sources are non-circular. The strong-uncorrelating transform is defined by a matrix $\mathbf{G}$ such that

$$\mathbf{G}\mathbf{R}_{XX}\mathbf{G}^H = \mathbf{I} \quad \text{and} \quad \mathbf{G}\mathbf{P}_{XX}\mathbf{G}^T = \widehat{\mathbf{\Lambda}}, \qquad (18)$$

where $\widehat{\mathbf{\Lambda}}$ is a diagonal real-valued matrix of ordered diagonal entries $1 \geq \widehat{\lambda}_1 \geq \widehat{\lambda}_2 \geq \cdots \geq \widehat{\lambda}_m \geq 0$. Methods for computing the strong-uncorrelating transform are given in [9, 10]. With this transformation, define

$$\mathbf{v}(k) = \mathbf{G}\mathbf{x}(k), \qquad (19)$$

as the prewhitened signal vector, such that

$$\mathbf{R}_{VV} = \frac{1}{N}\sum_{n=1}^{N}\mathbf{v}(n)\mathbf{v}^H(n) = \mathbf{I} \qquad (20)$$

$$\mathbf{P}_{VV} = \frac{1}{N}\sum_{n=1}^{N}\mathbf{v}(n)\mathbf{v}^T(n) = \widehat{\mathbf{\Lambda}}. \qquad (21)$$

Under prewhitening, $\mathbf{v}(k)$ and $\mathbf{s}(k)$ are related by

$$\mathbf{v}(k) = \mathbf{\Gamma}\mathbf{s}(k), \qquad (22)$$

where $\mathbf{\Gamma}$ is unitary ($\mathbf{\Gamma}\mathbf{\Gamma}^H = \mathbf{\Gamma}^T\mathbf{\Gamma}^* = \mathbf{I}$).

For single-source extraction, the source estimate is

$$y_t(k) = \mathbf{w}_t^T\mathbf{v}(k), \qquad (23)$$

where $\mathbf{w}_t$ is an m-dimensional vector of adjustable parameters at iteration t. The relationship between $\mathbf{w}_t$ and $\mathbf{c}_t$ is

$$\mathbf{c}_t = \mathbf{\Gamma}^T\mathbf{w}_t \qquad (24)$$

The second and fourth sample moments of $y_t(k)$ are

$$\frac{1}{N}\sum_{n=1}^{N}|y_t(n)|^2 = ||\mathbf{w}_t||^2 = ||\mathbf{c}_t||^2 \qquad (25)$$

$$\frac{1}{N}\sum_{n=1}^{N}|y_t(n)|^4 = \mathbf{c}_t^T\widehat{\mathbf{M}}_t\mathbf{c}_t^*, \qquad (26)$$

where we have defined the matrix

$$\widehat{\mathbf{M}}_t = \frac{1}{N}\sum_{n=1}^{N}\mathbf{s}(n)\mathbf{s}^H(n)\mathbf{c}_t^*\mathbf{c}_t^T\mathbf{s}(n)\mathbf{s}^H(n) \qquad (27)$$

The following theorem gives the structure of $\widehat{\mathbf{M}}_t$ [15]:

Theorem 5: *The limiting value of* $\mathbf{M}_t$ *is*

$$\lim_{N\to\infty}\widehat{\mathbf{M}}_t = \mathbf{c}_t^*\mathbf{c}_t^T + \mathbf{I}\mathbf{c}_t^H\mathbf{c}_t + \mathbf{\Lambda}\mathbf{c}_t\mathbf{c}_t^H\mathbf{\Lambda} + \mathbf{K}\mathrm{diag}\{\mathbf{c}_t\mathbf{c}_t^H\}, \qquad (28)$$

where the diagonal entries of diag$\{\mathbf{N}\}$ *are the diagonal elements of* $\mathbf{N}$.

Using this result, we can approximate

$$\widehat{\mathbf{M}}_t^T \mathbf{c}_t \approx \mathbf{K}\text{diag}\{\mathbf{c}_t\mathbf{c}_t^H\}\mathbf{c}_t + \mathbf{c}_t\left[2\mathbf{c}_t^H\mathbf{c}_t\right] + \mathbf{\Lambda}\mathbf{c}_t^*\left[\mathbf{c}_t^T\mathbf{\Lambda}\mathbf{c}_t\right]. \quad (29)$$

The first term is quite similar in form to (14), implying that the desired coefficient update before normalization is

$$\begin{aligned}\widetilde{\mathbf{c}}_t &\approx \mathbf{K}\text{diag}\{\mathbf{c}_t\mathbf{c}_t^H\}\mathbf{c}_t && (30)\\ &= \widehat{\mathbf{M}}_t^T\mathbf{c}_t - \mathbf{c}_t\left[2\mathbf{c}_t^H\mathbf{c}_t\right] - \mathbf{\Lambda}\mathbf{c}_t^*\left[\mathbf{c}_t^T\mathbf{\Lambda}\mathbf{c}_t\right]. && (31)\end{aligned}$$

Expressing this update in $\mathbf{w}_t$ coordinates gives

$$\begin{aligned}\widetilde{\mathbf{w}}_t = {} & \mathbf{\Gamma}^*\widehat{\mathbf{M}}_t^T\mathbf{\Gamma}^T\mathbf{w}_t - \mathbf{w}_t\left[2\mathbf{w}_t^T\mathbf{w}_t\right]\\ & - \mathbf{\Gamma}^*\mathbf{\Lambda}\mathbf{\Gamma}^H\mathbf{w}_t\left[\mathbf{w}_t^T\mathbf{\Gamma}\mathbf{\Lambda}\mathbf{\Gamma}^T\mathbf{w}_t\right]. \qquad (32)\end{aligned}$$

Finally, we notice that $\mathbf{\Gamma}\mathbf{\Lambda}\mathbf{\Gamma}^T \approx \mathbf{P}_{VV} = \widehat{\mathbf{\Lambda}}$ and

$$\mathbf{\Gamma}^*\widehat{\mathbf{M}}_t^T\mathbf{\Gamma}^T\mathbf{w}_t = \frac{1}{N}\sum_{n=1}^{N}|y_t(n)|^2 y_t(n)\mathbf{v}^*(n). \quad (33)$$

Combining the results gives the coefficient updates as

$$\begin{aligned}\widetilde{\mathbf{w}}_t &= \left(\frac{1}{N}\sum_{n=1}^{N}|y_t(n)|^2 y_t(n)\mathbf{v}^*(n)\right) - 2\mathbf{w}_t\\ &\quad - \widehat{\mathbf{\Lambda}}\mathbf{w}_t^*\left[\mathbf{w}_t^T\widehat{\mathbf{\Lambda}}\mathbf{w}_t\right] && (34)\\ \mathbf{w}_{t+1} &= \frac{\widetilde{\mathbf{w}}_t}{\sqrt{\widetilde{\mathbf{w}}_t^H\widetilde{\mathbf{w}}_t}} && (35)\end{aligned}$$

The above algorithm is similar to the FastICA algorithm for circular complex-valued sources in [8] for $G(y) = \frac{1}{2}y^2$. In fact, if $\widehat{\mathbf{\Lambda}} = \mathbf{0}$, the two updates are identical. The last term of (34), however, is novel, and it is critical to obtaining good performance for arbitrary mixtures.

The algorithm in (23), (34), and (35) relies on the strong-uncorrelating transform. Computing the strong-uncorrelating transform involves the Takagi factorization of a symmetric complex matrix, which generally requires specialized numerical code. For this reason, we offer an alternative implementation that employs ordinary prewhitening. In this version, find any prewhitening matrix $\widehat{\mathbf{G}}$ such that

$$\widehat{\mathbf{G}}\mathbf{R}_{XX}\widehat{\mathbf{G}}^H = \mathbf{I}. \quad (36)$$

and set $\mathbf{v}(k) = \widehat{\mathbf{G}}(k)\mathbf{x}(k)$, where

$$\widehat{\mathbf{P}} = \frac{1}{N}\sum_{n=1}^{N}\mathbf{v}(n)\mathbf{v}^T(n) = \widehat{\mathbf{G}}\mathbf{P}_{XX}\widehat{\mathbf{G}}^T. \quad (37)$$

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

It is possible to re-trace the steps taken to derive the updates in (34)–(35) under the assumption that $\widehat{\mathbf{P}}$ is not diagonal. The final version of the algorithm that results is

$$\begin{aligned}\widetilde{\mathbf{w}}_t &= \left(\frac{1}{N}\sum_{n=1}^{N}|y_t(n)|^2 y_t(n)\mathbf{v}^*(n)\right) - 2\mathbf{w}_t\\ &\quad - \widehat{\mathbf{P}}^*\mathbf{w}_t^*\left[\mathbf{w}_t^T\widehat{\mathbf{P}}\mathbf{w}_t\right] && (38)\\ \mathbf{w}_{t+1} &= \frac{\widetilde{\mathbf{w}}_t}{\sqrt{\widetilde{\mathbf{w}}_t^H\widetilde{\mathbf{w}}_t}} && (39)\end{aligned}$$

Comparing the updates in (38) and (34), 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 $y_t(n)$, $1 \leq n \leq N$ and the first term on the right-hand-side of (34) and (38).

The overall goal in our algorithm design was to obtain procedures for complex ICA that exhibit the fast, globally-convergent performance of the real-valued FastICA algorithm. Do (34)–(35) and (38)–(39) achieve this end? The following theorem indicates that the answer is yes [15].

Theorem 6: *As $N \to \infty$, both (34)–(35) and (38)–(39) 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 $\{e^{j\theta_i}[\text{sgn}(\kappa_i)]^t\}$, $\mathbf{a}_t$ is a positive-valued m-dimensional vector obeying the relationships*

$$\widetilde{\mathbf{a}}_t = \mathbf{K}_a\mathbf{F}(\mathbf{a}_t), \qquad \mathbf{a}_{t+1} = \frac{\widetilde{\mathbf{a}}_t}{\sqrt{\widetilde{\mathbf{a}}_t^T\widetilde{\mathbf{a}}_t}} \quad (40)$$

where the diagonal matrix $\mathbf{K}_a$ has diagonal entries $\{|\kappa_1|, \ldots, |\kappa_m|\}$ 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, their convergence performance is identical to that of the real-valued FastICA algorithm with kurtosis contrast, where complex-source kurtoses replace real source kurtoses and coefficient amplitudes replace coefficient values in the evolutionary behavior.

Our algorithms also do not change the output signal phase during their operation except for a sign flip during odd-valued iterations, a desirable attribute for some applications.

5. FIXED-POINT ALGORITHMS FOR SEPARATING ARBITRARY COMPLEX MIXTURES

To extend either of our algorithms to general m-source extraction, we use similar concepts as in the real-valued FastICA algorithm extended to the complex realm. Since $\mathbf{v}(k)$ is related to $\mathbf{s}(k)$ through the unitary matrix $\mathbf{\Gamma}$, all m sources can be extracted by applying m versions of either algorithm to $\mathbf{v}(k)$ and constraining the resulting $\mathbf{w}_{it}$ vectors to be unitary. Unitarity could be maintained either

1. sequentially through a Gram-Schmidt or QR method, or

```
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
    Wold = zeros(m,1);
    Wt = W(:,i);        k=0;
    while (abs(abs(Wold'*Wt)-1)>1e-4)*(k<100)
        Wold = Wt;     k=k+1;
        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;
```

Table 1: A MATLAB implementation of the fixed-point algorithm that uses sequential orthogonalization.

2. jointly through a symmetric unitarization procedure using an inverse matrix square root or an adaptive method.

Sequential unitarization procedures work without knowledge of the number of sources in the mixture and are provably-convergent as $N \rightarrow \infty$, but sources extracted later in the procedure contain greater amounts of interference. Joint unitarization procedures often perform better but do not work as well when Gaussian sources are present and are not guaranteed to converge for $m > 2$. It is suggested that one alternates between sequential and joint unitarization procedures to obtain best performance.

Table 1 lists the CFPA1 algorithm, a sequential implementation of m versions of (38)–(39) with Gram-Schmidt orthogonalization using the MATLAB technical computing environment. Table 2 provides the CFPA2 algorithm, a parallel implementation of m versions of (38)–(39) in which joint unitarization is used. Versions employing (34)–(35) have been omitted but are simple to construct given Takagi factorization software.

6. SIMULATIONS

We now explore performance behaviors of algorithms via simulations. All evaluations are performed on synthetic data using the MATLAB technical computing environment. We have used the average inter-channel interference (ICI) to measure separation performance, which for the combined system matrix $\mathbf{C}_t = \mathbf{W}_t\widehat{\mathbf{G}}\mathbf{A}$ with (i,j)th element c_{ijt} is given by

$$ICI_t = \frac{1}{m}\sum_{i=1}^{m}\left(\frac{\sum_{l=1}^{m}|c_{ilt}|^2 - \max_{1\leq k\leq m}|c_{ikt}|^2}{\max_{1\leq k\leq m}|c_{ikt}|^2}\right) \quad (41)$$

This performance measure does not measure whether all sources are extracted uniquely, although the algorithms being compared extract different sources due to coefficient orthogonality. The mixing matrix $\mathbf{A}$ has been generated randomly for each simulation run using an SVD-like combination of two random complex orthogonal matrices and a set of complex diagonal elements whose amplitudes were restricted to the interval $[0.2, 1]$. Both noiseless and noisy mixtures have been used, in which additive circular uncorrelated Gaussian noises with variances $\sigma_\nu^2 = 0.001$ was used as the measurement interference. Our study of noisy signal mixtures is to make sure that the various algorithms are robust to low-level additive sensor noise.

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

[N,m] = size(x);   k=0;
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);
Wold = zeros(m);   D = W;
while(norm(abs(Wold'*W)-eye(m),'fro')>(m*1e-4))*(k<15*m)
    Wold = W;      k=k+1;
    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
B = Ghat*W;        y = v*W;
```

Table 2: A MATLAB implementation of the fixed-point algorithm that uses joint symmetric orthogonalization.

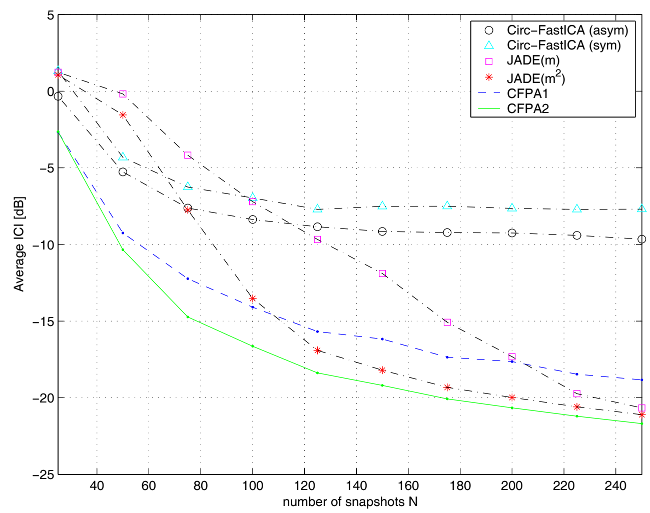

Fig. 1: Average ICI as a function of data record length N for a noiseless six-source demixing task.

We compare the separation performance of our algorithms to two different versions of two well-known existing methods for complex ICA: JADE [1], and the complex FastICA algorithm in [8] that assumes circularly-symmetric source distributions with $G(|y|^2) = 0.5|y|^2$. The two JADE algorithms used simultaneous diagonalization of m and m^2 cumulant matrices, respectively, and the two versions of the FastICA algorithm in [8] employed symmetric and asymmetric deflation procedures, respectively. One thousand eval-

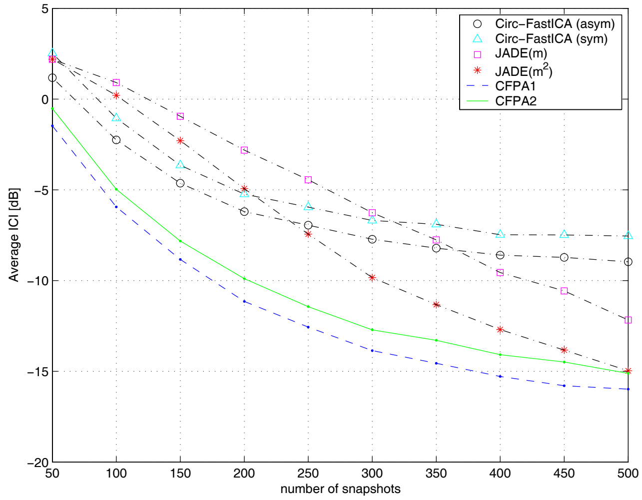

Fig. 2: Average ICI as a function of data record length N for a noisy seven-source demixing task.

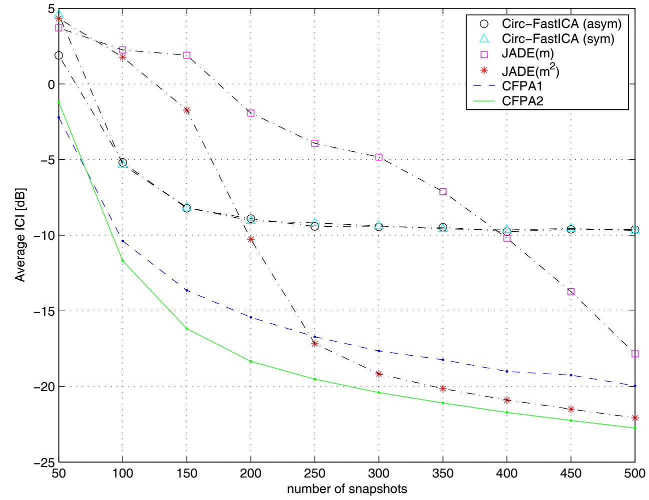

Fig. 3: Average ICI as a function of data record length N for a more-challenging noiseless nine-source demixing task.

uations of each method have been used to determine the averaged performance statistics shown.

Consider noiseless six-source mixtures of two real-valued binary-$\{\pm 1\}$ sources, two 4QAM sources, and two 16QAM sources. Fig. 1 shows the average ICI of the six algorithms for different data block lengths N. Our proposed methods perform better than either version of JADE and either version of the algorithm in [8] for small sample sizes, offering separation of between 12.5 and 15 dB for $N = 75$ snapshots. The complex FastICA procedure for circular sources in [8] produces a biased result. The two JADE algorithms perform better than CFPA1 for larger block lengths. CFPA2 performs the best for all block lengths.

Fig. 2 shows the average ICI of the six algorithms for a seven-source noisy mixture, in which a circularly-symmetric complex Gaussian source has been included in the above signal set with additive Gaussian sensor noise. In this case, our proposed symmetric-deflation-based technique performs the best for all block lengths considered. Additional examples with nine- and ten-source mixtures are given in [15].

We now consider a scenario in which three distributions – Uniform-$[-\sqrt{3}, \sqrt{3}]$, unit-variance Laplacian, and binary – are used to generate nine sources by (a) taking all possible distribution pairs to create the real and imaginary parts of six complex sources, and (b) adding three additional real-valued sources of each distribution. Shown in Fig. 3 are the behaviors of the six algorithms in this case. The proposed methods are superior for $N \leq 250$, and CFPA2 performs the best for all block lengths.

7. CONCLUSIONS

In this paper, we have presented algorithms for separating mixtures of independent, non-circularly-symmetric, and non-Gaussian sources. Our algorithms 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 show that the techniques have finite-sample separation performance that often meets or exceeds that of existing approaches for complex-valued blind source separation, especially for small record lengths.

Acknowledgement: The author would like to thank Jan Eriksson and Visa Koivunen for various helpful discussions.

8. REFERENCES

[1] J.-F. Cardoso and A. Soloumiac, "Blind beamforming for non-Gaussian signals," *IEE Proc. F,* vol. 140, pp. 362-370, Dec. 1993.

[2] P. Comon, "Independent component analysis: A new concept?" *Signal Processing,* vol. 36, pp. 287-314, Apr. 1994.

[3] A. Hyvärinen and E. Oja, "A fast fixed-point algorithm for independent component analysis," *Neural Computation,* vol. 9, pp. 1483-1492, Oct. 1997.

[4] S. Amari, S.C. Douglas, A. Cichocki, and H.H. Yang, "Multichannel blind deconvolution using the natural gradient," *Proc. 1st IEEE Workshop Signal Proc. Adv. Wireless Commun.,* Paris, France, pp. 101-104, Apr. 1997.

[5] V. D. Calhoun, T. Adali, G. D. Pearlson, P. C. van Zijl, and J. J. Pekar, "Independent component analysis of fMRI data in the complex domain," *Magn Reson. Med.,* vol. 48, pp. 180-192, 2002.

[6] J. Anemuller, T. J. Sejnowski, and S. Makeig, "Complex independent component analysis of frequency-domain electroencephalographic data," *Neural Networks,* vol. 16, pp. 1311–1323, 2003.

[7] M. Martone, "Fast adaptive super-exponential multistage beamforming for cellular basestation transceivers with antenna arrays," *IEEE Trans. Vehicular Tech.,* vol. 48, no. 4, pp. 1017-1028, July 1999.

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

[9] J. Eriksson and V. Koivunen, "Complex-valued ICA using second order statistics," *Proc. IEEE Workshop on Machine Learning Signal Processing,* Sao Luis, Brazil, pp. 183-191, October 2004.

[10] J. Eriksson and V. Koivunen, "Complex random vectors and ICA models: Identifiability, uniqueness, and separability," *IEEE Trans. Inform.Theory,* in press.

[11] O. Shalvi and E. Weinstein, "Super-exponential methods for blind deconvolution," *IEEE Trans. Inform. Theory,* vol. 39, pp. 504-519, Mar. 1993.

[12] S.C. Douglas, "On the convergence behavior of the FastICA algorithm," *Proc. Fourth Symp. Indep. Compon. Anal. Blind Signal Separation,* Kyoto, Japan, pp. 409-414, Apr. 2003.

[13] S.C. Douglas, "A statistical convergence analysis of the FastICA algorithm for two-source mixtures," *Proc. 39th Asilomar Conf. Signals, Syst., Comput.,* Pacific Grove, CA, Oct. 2005 (in press).

[14] S.C. Douglas, Z. Yuan, and E. Oja, "Average convergence behavior of the FastICA algorithm for blind source separation," *Proc. 6th Int. Conf. Indep. Compon. Anal. Blind Signal Separation,* Charleston, SC, Mar. 2006 (in press).

[15] S.C. Douglas, " Fixed-point algorithms for the blind separation of arbitrary complex-valued non-Gaussian signal mixtures," submitted.