Adaptive Complex-Valued Independent Component Analysis Based on Second-Order Statistics

Yanfei Jia\textsuperscript{1,2} and Xiaodong Yang\textsuperscript{1,3}

\textsuperscript{1}College of Information and Communication Engineering, Harbin Engineering University, Heilongjiang 150001, China
\textsuperscript{2}College of Electrical and Information Engineering, Beihua University, Jilin 132012, China
\textsuperscript{3}Collaborative Research Center, Meisi University, Tokyo 1918506, Japan

Correspondence should be addressed to Yanfei Jia; jia.yanfei@163.com

Received 1 April 2016; Revised 28 July 2016; Accepted 16 August 2016

1. Introduction

Blind source separation (BSS) is the separating of a set of source signals from a set of mixed signals without the aid of information (or with very little information) about either the source signals or the mixing process. Independent component analysis (ICA) is an attractive approach for solving blind source separation problems. ICA can be divided into real-valued ICA and complex-valued ICA according to the mixed signals. Complex-valued ICA is widely used to estimate the mixing matrix or to separate complex-valued mixed signals, such as frequency domain signals [1, 2], digital communication signals [3, 4], functional magnetic resonance imaging signals [5], and power system signals [6].

Studies of complex-valued ICA can be divided into three categories. The first category includes methods based on a nonlinear function, such as complex-valued fastICA (C-fastICA) [7], noncircular complex fastICA (NC-fastICA) [8], complex maximization of non-Gaussianity (CMN) [9], complex-valued ICA by entropy bound minimization (CEBM) [10], complex-valued ICA by entropy rate bound minimization (CERBM) [11], and others [3, 12]. The second category includes methods that are based on kurtosis or higher-order cumulants, such as joint approximative diagonalization of eigenmatrix (JADE) [4], kurtosis maximization (KM) [13], pseudo-Euclidean gradient iteration ICA (GEGI-ICA) [14], and others [15–17]. The third category includes methods based on second-order statistics, such as strong-uncorrelating transform (SUT) [18, 19] and its adaptive algorithms [2, 20–22] and pseudo-uncorrelating transform (PUT) [23]. Recently, the performance and separability of complex-valued Gaussian mixtures of SUT method have also been studied [24, 25]. Every complex-valued ICA category has its own merits and appropriate application conditions. The methods based on second-order statistics have a simple structure and low computation complexity and are suitable for complex Gaussian and non-Gaussian noncircular signals. In contrast, ICA methods in the first and second categories are not suitable for use with complex Gaussian noncircular signals.

The major advantage of SUT is that “whenever applicable, remains perhaps the simplest and most accessible approach” [24]. SUT is a batch algorithm and cannot be used to process signals in real time, so some adaptive complex-valued ICA algorithms have been proposed based on second-order statistics [2, 20–22]. Compared with other complex-valued ICA
strategies, adaptive complex-valued ICA algorithms based on second-order statistics are simpler in structure and do not require the probability density of the real and imaginary parts of a complex-valued source signal to be non-Gaussian. The Scott method [20] proposes an updating formula of the separating matrix for adaptive complex-valued ICA without mathematical speculation. The Cong method [2] simultaneously uses diagonal covariance and pseudocovariance noncircular signals as the cost function to deduce the adaptive complex ICA. The convergence condition of the Scott and Cong methods requires that the covariance and pseudocovariance of the separated signal are simultaneously diagonal. For example, if only the covariance of the separated signal is diagonal, the method is unable to reach convergence until the pseudocovariance is also diagonal. This requirement could affect convergence speed. The Yang method [22] uses a two-step serial updating method to make the separated signals satisfy the above convergence condition. In the second step, Yang uses the orthogonal method to force the separating matrix to be a unitary matrix. This changes the updating direction of the separating matrix and leads to slow convergence speed.

To increase the rate of convergence, a fast complex-valued ICA method is proposed in this work. The proposed method first extends the real-valued whitening process to a complex-valued domain to provide unit variance for the processed signal. Second, this work uses the restriction that the pseudocovariance matrix of the separated signals is a diagonal matrix to construct cost function and optimize the cost function using the geodesic method. This avoids computing the square root and inverse of the separating matrix and also keeps the separating matrix to be an orthogonal matrix, without any forcing operation. This improves the convergence speed of the proposed method compared to the other adaptive methods.

2. Complex-Valued ICA and Second-Order Statistics

2.1. Complex-Valued Linear ICA Model. Generally, a linear complex-valued ICA model that is noise-free can be expressed as follows:

\[
x = As,
\]

where \( s = [s_1, s_2, \ldots, s_n]^T \) is the unknown column vector of source signals, \( n \) is the number of source signals, \( A \) is the unknown complex-valued mixing matrix, \( x = [x_1, x_2, \ldots, x_m]^T \) is the column vector of observed complex-valued mixed signals, and \( m \) is the number of observed signals. The components of the source signals are mutually independent. Most complex-valued ICA algorithms assume that the number of observed signals is not less than the source signals, and only one Gaussian source signal is allowed. The aim of complex-valued ICA is to search the separating matrix and estimate the source signals and mixing matrix. Given that complex-valued ICA does not utilize any information about the source signals or mixing matrix, it has some indeterminacy in amplitude, sequence, and phase. This indeterminacy does not affect the shape of the estimated source signal waveform, which contains most information about source signals.

2.2. Second-Order Statistics of Complex-Valued Signals. Assume a complex-valued random column vector \( s = s_R + j s_I \), where \( s_R \) and \( s_I \) are the real and imaginary part of \( s \), respectively, and \( j = \sqrt{-1} \). The expectation \( E[\cdot] \) of the random vector \( s \) is defined as follows:

\[
E[s] = E[s_R] + jE[s_I].
\]

Its covariance matrix \( \text{cov}(s) \) is defined as follows:

\[
\text{cov}(s) = E[(s - E(s))(s - E(s))^H],
\]

where \((\cdot)^H\) denotes the Hermitian transpose. Its corresponding pseudocovariance matrix is defined as follows:

\[
p\text{cov}(s) = E[(s - E(s))(s - E(s))^T],
\]

where \((\cdot)^T\) denotes the transpose. The covariance matrix together with the pseudocovariance matrix is the full expression of second-order statistics [19]. If the pseudocovariance matrix equals zero, the random vector is considered circular or proper. If both the covariance matrix and pseudocovariance matrix of the random vector are diagonal with nonzero diagonal elements, the random vector is noncircular or improper, and components of the random vector are called strong uncorrelated components.

2.3. Complex-Valued ICA Based on SUT. For any complex random vector \( x \), if the vector can be transformed into a random vector \( s \) by use of a nonsingular square matrix \( w \), where \( s = [s_1, s_2, \ldots, s_n]^T = wx \) has covariance that is a unit matrix and pseudocovariance that is a diagonal matrix with diagonal elements between zero and one, then the matrix \( w \) is called SUT. If the observed signal is the complex random vector \( x \) and the source signal is \( s \), then the SUT is the separating matrix in complex-valued ICA. The procedure for complex-valued ICA based on SUT is as follows [18].

1. Whitening the complex-valued observed signals \( x \): the whitening procedure is given by

\[
z = Bx = \text{cov}(x)^{-1/2} x,
\]

where the whitening matrix \( B \) is the inverse of the matrix square root of the covariance matrix and \( z \) is the whitened signal with a unit covariance matrix.

2. Determining the separating matrix of the whitened signal by use of Takagi’s factorization: this is done according to

\[
p \text{cov}(z) = U\Lambda U^T.
\]

From (5) and (6) we obtain the separating matrix \( w = U^H B \).
3. Proposed Adaptive Complex-Valued ICA

In this section, we describe an adaptive fast convergence complex-valued ICA algorithm based on second-order statistics, used in the SUT method. This is unlike other adaptive complex-valued ICA methods that simultaneously force separated signals to comply with second-order statistics. Instead, this method uses an adaptive serial updating method to realize the SUT. First, we use an adaptive method to whiten the observed signals. The cost function used in real-value ICAs can be extended to the CVA.

To avoid computing the matrix inverse, a complex-valued natural gradient is used to simplify (8):

$$J(w) = -\frac{1}{2} \left[ \log(\det(uuw^H)) - \sum_{i=1}^{n} E\{|y_i|^2\} \right], \quad (7)$$

where $w$ is the whitening matrix and $y_i$ is the $i$th whitening signal. In complex-valued signal processing, the steepest descent direction of cost function (7) is

$$\frac{\partial J(w)}{\partial \omega^*} = -\frac{1}{2} \left( \frac{\partial \log(\det(uuw^H))}{\partial \omega^*} - \sum_{i=1}^{n} E\{|y_i|^2\} \right) \quad \text{(8)}$$

where $x$ is the observed signal, $y = wx$, $\partial \omega^*/\partial \omega^* = 1$, and $\partial \omega^*/\partial \omega^* = 0$. To avoid computing the matrix inverse, a complex-valued natural gradient is used to simplify (8):

$$\Delta w = \frac{\partial J(w)}{\partial \omega^*} (w^H w) = \left[ -\left(wuH\right)^{-1} w + E\left(yx^H\right) \right] (w^H w) = E\left(yy^H\right) w - w = E\left(yy^H\right) - I \right] w. \quad \text{(9)}$$

So, adaptive whitening can be expressed as follows:

$$w(k+1) = w(k) + \mu \left[I - E\left(yy^H\right)\right] w(k). \quad \text{(10)}$$

If we use the instantaneous value instead of the expected value in (10), we obtain the adaptive real-time whitening method:

$$w(k+1) = w(k) + \mu \left[I - yy^H\right] w(k). \quad \text{(11)}$$

Second, we must modify the separated signals to satisfy a diagonal pseudocovariance matrix while keeping the covariance matrix as a unit matrix. We use the cost function in [22], which can be expressed as follows:

$$J_2(v) = \frac{1}{2} \left\| \Lambda - E\left[zz^T\right] \right\|^2, \quad \text{(12)}$$

where $z = yy$, $v$ is the separating matrix of the whitened signals and $\Lambda$ is the diagonal matrix of $E[zz^T]$. The ordinary gradient with $v^*$ is as follows:

$$\frac{\partial J_2(v)}{\partial v^*} = \frac{1}{2} \left[ \frac{\partial}{\partial v^*} \left( \Lambda - E\left[zz^T\right] \right) \right]^2$$

$$= \frac{1}{2} \left[ \text{tr}\left(\left(\Lambda - E\left[zz^T\right]\right)\left(\Lambda - vE\left[yy^T\right]\right)^H\right) \right]$$

$$= \frac{1}{2} \left[ \text{tr}\left(\left(\Lambda - vE\left[yy^T\right]\right)\left(\Lambda - vE\left[yy^T\right]\right)^H\right) \right]$$

$$= -\left(\Lambda - vE\left[yy^T\right]\right) E\left[v^* v^* + 1\right]. \quad \text{(13)}$$

The update of $v$ can be written as follows:

$$v(k+1) = v(k) + \mu \left(\Lambda - vE\left[yy^T\right]\right) E\left[v(k)^* P^*\right], \quad \text{(14)}$$

where $P = E[yy^T]$ is the correlated matrix of the whitened signal. At the convergence point, the pseudocovariance matrix of the separated signal is diagonal. To keep the covariance matrix of the separated signal as a unit matrix, the separating matrix $v$ must be a unitary matrix. In [22], they directly used the method of fixed-point fastICA to force the separating matrix to be a unitary matrix:

$$v(k+1) = v(k) + \mu \left(\Lambda - vE\left[yy^T\right]\right) E\left[v(k)^* P^*\right], \quad \text{(15)}$$

This approach has two major drawbacks. One is that (16) changes the steepest gradient direction in every iteration, which slows the convergence speed. The second is that (16) must compute the square root and the inverse of the separating matrix in every iteration, which increases the algorithm computation complexity, slowing the time of convergence.

To overcome this problem, we use a geodesic method to search the optimized separating matrix $v$. The geodesic method causes the separating matrix to move on the surface of the orthogonal matrix to converge to a local minimum without a forcing operation. The geodesic method is given by

$$v(k+1) = v(k) + \mu \left(\Lambda - v(k) E\left[yy^T\right]\right) E\left[v(k)^* P^*\right], \quad \text{(16)}$$

where $P = \exp\left[u\left(\Delta v(k+1) + v(k)^* P^*\right)\right]. \quad \text{(17)}$

If $v(k)$ is a unitary matrix, then $v(k+1)$ is also a unitary matrix. By using the geodesic method, we do not need additional operations to make the separating matrix an orthogonal matrix and change its search direction.

Using the geodesic method with self-tuning [26] to optimize the cost function (12), we can describe a fast convergence complex-valued ICA method. The implementation process of the proposed adaptive ICA method is as follows:

1. Initialize the whitening matrix and separating matrix using unit matrix, learning rate $\mu_1$, and $\mu_2$, and iterative number for optimizing (7) and (12), respectively.
(2) Use (10) to whiten the observed signal $x$ and obtain the whitening signal $y = wx$ and whitening matrix $w$.

(3) Compute the gradient of the cost function in Riemannian space, which can be expressed as follows:

$$G(k) = (\Lambda - R)E[F^H] - E[F] (\Lambda - R^H),$$

where $\Lambda$ is a diagonal matrix with diagonal elements $R$, $R = \nu(k)Pv(k)^T$, $P = E[yy^T]$, and $F = \nu(k)^*P^H$.

(4) Compute the rotation matrix $D(k) = \exp(-\mu_2 G(k))$ and $Q(k) = D(k)D(k)^T$.

(5) If $J \geq (\mu_2/2)$, $\mu_2 = 2\mu_2$, where

$$J = \|\Lambda - R\|^2 - \|\Lambda - Q(k)RQ(k)^T\|^2,$$

where $\Lambda$ and $\Lambda_2$ are diagonal matrices corresponding to $R$ and $Q(k)RQ(k)^T$, respectively.

(6) If $J < (\mu_2/2)$, $\mu_2 = 0.5\mu_2$.

(7) Update the separating matrix

$$v(k + 1) = \exp(-\mu_2 G(k))v(k).$$

(8) If real[trace($G(k)G(k)^H$)] is sufficiently small, then STOP; else return to step (3).

4. Experimental Results and Analysis

In order to test the algorithm, we used five synthesized signals with different spectral coefficients, three digital communication signals with different spectral coefficients, and three synthesized signals of which two signals have same spectral coefficients as the source signals. For simplicity, we directly used the expectation of the signal instead of the instantaneous value. Quality of separation was assessed using the performance index (PI), a widely used index in ICA. PI can be expressed as [27]

$$\text{PI}(H) = \frac{1}{n(n-1)} \left\{ \sum_{i=1}^{n} \left( \sum_{j=1}^{n} \frac{|h_{ij}|}{\max_i |h_{ij}|} - 1 \right) + \sum_{j=1}^{n} \left( \sum_{i=1}^{n} \frac{|h_{ij}|}{\max_j |h_{ij}|} - 1 \right) \right\},$$

where $h_{ij}$ is the $(i, j)$ element of the global system matrix $H = \omega BA$, $\omega B$ is the separating matrix of mixed signal, $A$ is the mixing matrix, and $\max_i |h_{ij}|$ and $\max_j |h_{ij}|$ are the maximum absolute value of the elements in the $i$ row and $j$ column vector $H$, respectively. When perfect separation is achieved, the performance index is zero. “In practice, the value of performance index $10^{-2}$ gives quite a good performance” [27]. The smaller the value of PI, the better the performance.
The mixing matrix is a complex-valued random matrix with real and imaginary parts generated from a random uniform distribution between 0 and 1. All algorithms have the same learning rate of 0.01 and were run 100 times. Each time, the source signal and mixing matrix was independently generated.

In contrast, convergence curves are shown in Figure 1 that correspond to the four methods: Yang method [22], Scott method [20], SUT method [18], and our proposed method. Every method has 100 convergence curves, and every convergence curve corresponds to results from one run. The SUT method is a batch method without iterative computations. Therefore, the convergence curves are straight lines. From Figure 1, we can see that all the convergence curves of the proposed method are more closer than the other adaptive methods except for the SUT method. This suggests that the proposed method shows improved, stable performance for different mixed sources that is better than the other adaptive methods. The SUT method shows the smallest fluctuation range, followed by the proposed method, Scott method, and then the Yang method. This indicates that the proposed method is more suitable for processing different mixed signals than the other adaptive methods, except for the SUT method. Although the performance of SUT is more stable than the other methods for separating different mixed signals, its realization involves Takagi’s factorization that is difficult to implement and is not suitable for real-time separation of mixed signals. The adaptive complex-valued BSS method is easy to perform and is more appropriate for real-time separation of mixed signals.

Average convergence curves for the four methods are shown in Figure 2. From Figure 2, we see that the Yang method does not converge to a stationary point until 40000 iterations; the Scott method starts to converge close to 35000 iterations; the proposed method starts to converge after about 2000 iterations. Thus, the proposed method has a faster convergence speed. The performance index is larger than the proposed method when the Scott method converges to a stationary point. This indicates that the proposed method has a smaller error than the Scott method. The performance

**Figure 3: Original signals, mixed signals, and separated signals in the digital communication system.**
indices of the proposed method and the SUT method are very similar, indicating that the two methods have almost the same amount of error.

In the second experiment, we supposed that three digital communication signals (8QAM, 4QAM, and BPSK) impinge on a uniform linear antenna array with three elements from directions of 10°, 25°, and 70°. In Figure 3, the first row gives the original source signals, the second row gives the three mixed signals that are separately received by the three elements of antenna, and the third row provides the separated signals obtained using the proposed method. Comparing the source signals with the separated signals, we see that the constellation of separated signals is almost the same as the source signals, except the sequence, amplitude, and phase, which are inherently indeterminate. This shows that the proposed method is valid for the supposed communication signals.

The average convergence curves for the four methods are shown in Figure 4 from an average of 100 different simulation runs with a learning rate of 0.01. From Figure 4, we see that the proposed method starts to converge after 150 iterations, the Scott method starts to converge after 4500 iterations, and the slowest to converge is the Yang method, which starts to converge after 17000 iterations. Thus, the proposed method has faster convergence than the other adaptive methods. When the proposed method converges to the stationary point, the performance index curve of proposed method and SUT method are the same. This means that the two methods have the same error for the communicating signals.

In the third experiment, three random complex-valued signals were used as source signals, with spectral coefficients of 0, 0.6, and 0.6. Their imaginary and real parts were generated by a random uniform distribution function. Average convergence curves from an average of 100 different simulation runs with a learning rate of 0.01 are shown for the four methods in Figure 5. From Figure 5, we see that the performance indexes of Yang method, the Scott method, and the proposed method are less than 0.1 at the stationary point. The average performance index of SUT is about 0.33, which is far greater than 0.1. According to [27], this means that the three adaptive methods successfully separated the mixed signals but the SUT method failed for the mixed signals. The SUT method includes Takagi’s factorization to factorize the pseudocovariance matrix. Therefore, it is not suitable for noncircular signals with the same spectral coefficients.

The proposed method has two stages. The convergence curves shown in all figures are the convergence curves only for the second stage. For the first stage, the whitening signal converges to the unit matrix in first experiment after about 600 iterations and after about 100 iterations in the second and third experiments. Compared with other methods, the total iterations required for the proposed method are far less than other methods.

5. Conclusions

This paper proposes an adaptive complex-valued ICA method for noncircular signals based on second-order statistics and the geodesic method. The proposed method has faster convergence and smaller error than the other adaptive methods. For different mixing source signals, the proposed method has better performance and faster convergence than the Scott method. For source signals with different spectral coefficients, the proposed method and the SUT method have almost the same error. However, the SUT method is not suitable for source signals that some of source signals have the same spectral coefficients.

Competing Interests

The authors declare that they have no competing interests.
Acknowledgments

This work was supported by the National Natural Science Foundation of China (61271115) and the Foundation of Jilin Educational Committee (2015235).

References


Submit your manuscripts at http://www.hindawi.com