Published in final edited form as:

J Stat Comput Simul. 2015 April ; 85(6): 1151–1164. doi:10.1080/00949655.2013.867961.

Evaluating Model Misspecification in Independent Component Analysis

Seonjoo Lee^{a,*}, Brian S. Caffo^b, Balaji Lakshmanan^c, and Dzung L. Pham^d

^a Department of Psychiatry and Biostatistics, Columbia University

^b Department of Biostatistics, Johns Hopkins University

^c Kennedy Krieger Institute

^d Center for Neuroscience and Regenerative Medicine, The Henry M. Jackson Foundation

Abstract

Independent component analysis (ICA) is a popular blind source separation technique used in many scientific disciplines. Current ICA approaches have focused on developing efficient algorithms under specific ICA models, such as instantaneous or convolutive mixing conditions, intrinsically assuming temporal independence or autocorrelation of the sources. In practice, the true model is not known and different ICA algorithms can produce very different results. Although it is critical to choose an ICA model, there has not been enough research done on evaluating mixing models and assumptions, and how the associated algorithms may perform under different scenarios. In this paper, we investigate the performance of multiple ICA algorithms under various mixing conditions. We also propose a convolutive ICA algorithm for echoic mixing cases. Our simulation studies show that the performance of ICA algorithms is highly dependent on mixing conditions and temporal independence of the sources. Most instantaneous ICA algorithms fail to separate autocorrelated sources, while convolutive ICA algorithms depend highly on the model specification and approximation accuracy of unmixing filters.

Keywords

Convolutive Mixing; Independent Component Analysis; Model Misspecification

1. Introduction

Independent component analysis (ICA) aims to identify an unknown system driven by N latent inputs, based on M outputs assuming only the statistical independence of the underlying sources. It is of importance in many biomedical applications, such as electroencephalography (EEG) and magnetoencephalography (MEG). For example, see [1, 2]. The literature has considered mainly two linear mixing conditions: instantaneous and convolutive linear mixings. However, model evaluation between instantaneous and

*Corresponding author. seonjool@gmail.com.

convolutive ICA models have not been studied yet. Thus, in this paper, we study the effects of model specification in ICA and propose a method to guide model identification.

In instantaneous mixing cases, the observations $\mathbf{X}$ can be expressed as a weighted sum of the sources $\mathbf{S}$:

$$\mathbf{X}(t) = \mathbf{A}\mathbf{S}(t), \quad t=1, \dots, T, \quad (1)$$

where $\mathbf{X}(t)$ and $\mathbf{S}(t)$ are vectors of dimension M and N , respectively, and $\mathbf{A}$ is an $M \times N$ mixing coefficient matrix. For simplicity, we focus on the case $M = N$ for the rest of the manuscript. Many algorithms are available for instantaneous mixtures based on different independence measurements, such as high-order statistics [3], information theoretic measurements [4-6], canonical correlations in a reproducing kernel Hilbert space [7], maximum likelihood [8-11], characteristic function [12, 13] and the Whittle likelihood [14].

In the cases where the mixing process yields convolutions with time delays, the relations between the sources and the observation can be expressed as

$$\mathbf{X}(t) = \sum_{k=0}^P \mathbf{A}_k \mathbf{S}(t-k), \quad t=1, \dots, T, \quad (2)$$

where P is large enough that all correlations in the process $\mathbf{X}(t)$ can be explained by the mixing process. As a special case of convolutive mixing, [15] and [16] focused on anechoic mixing cases, where the sources are mixed with only time delays, that is, without echoes. For general convolutive mixing, most methods are based on a joint diagonalization algorithm. For white noise sources, methods include the joint diagonalization of the covariance matrix or second order statistics on spectral domain [17] and singular value decompositions (SVD) on the power spectrum and the higher order poly spectra [18]. Non-white sources were dealt with by [19]. It is further generalized to nonlinear process sources using higher-order spectra [20]. Also, multivariate (vector) autoregressive moving-average based methods have been proposed [21-24]. For further review, see [25] and chapter 8 in [26] and references therein.

The ICA literature shows that all ICA algorithms have been built on one of four mixing conditions: 1) instantaneous mixing of temporally independent sources, 2) instantaneous mixing of temporally autocorrelated sources, 3) convolutive mixing of temporally independent sources, and 4) convolutive mixing of temporally autocorrelated sources. Although identifying mixing models is critical, our literature review showed little research regarding the effects of misspecified ICA models, excepting cases of instantaneous mixing. For example, [14] has indicated that most marginal independence measure based approaches, which intrinsically assume temporal independence of sources, fail to separate autocorrelated sources. Model identification becomes even more difficult in convolutive mixing cases without proper constraints when the sources are autocorrelated. Thus, most currently existing convolutive ICA algorithms assume the sources are not autocorrelated.

In this manuscript, we investigate performance of instantaneous and convolutive ICA algorithms under different mixing conditions. For instantaneous ICA, there are many

publicly available algorithms, while there are few publicly available convolutive ICA algorithms. For instantaneous ICA algorithms, we consider marginal independence measure-based methods, such as JADE [3] and fastICA [6] and autocorrelation modeling-based methods, such as cICA-YW [14]. The first two ICA algorithms are popular marginal independence ICA approaches based on information theoretic optimization. The third algorithm is based on the Whittle likelihood and has demonstrated better performance for auto-correlated sources.

Due to the lack of publicly available convolutive ICA algorithms and for simplicity of comparison, we consider echoic mixing cases for the convolutive mixing condition, where the sources are mixed with decaying echoes. We propose a convolutive ICA method specifically for this case. The proposed algorithm is based on the second moments of the latent sources to estimate echoic mixing matrices. The main idea of our method lies in utilizing autocorrelation structures of convolutive mixtures.

The rest of manuscript is laid out as follows: Section 2 describes the model specification problem in ICA and a newly proposed convolutive ICA method under a simple convolutive setting in Section 3. In Section 4, we investigate the effects of model misspecification via extensive simulation studies and analyze EEG data and compare various ICA algorithms.

2. Model Specification in Independent Component Analysis Models

In this section, we discuss the effects of model mis-specification in instantaneous and convolutive ICA. The literature has shown that the instantaneous ICA algorithms based on marginal independence are not able to separate autocorrelated sources. For example, maximum likelihood based or entropy based ICA algorithms assume that the joint density of the source $\mathbf{S}(t)$ can be expressed as a product of the marginals:

$$p(S_1, \dots, S_M) = \prod_{t=1}^T \prod_{j=1}^M p(S_j(t)).$$

For the autocorrelated sources, the equality does not hold, and this leads inaccurate estimation of ICA algorithms.

For convolutive mixing cases described in (2), the instantaneous ICA algorithms cannot separate the mixtures. That is, we cannot find a unmixing matrix $\mathbf{W}$ such that

$\mathbf{W}\mathbf{X}(t) = \sum_{k=0}^P \mathbf{W}\mathbf{A}_k \mathbf{S}(t-k)$, where $\mathbf{W}\mathbf{A}_0 = \mathbf{I}$ and $\mathbf{W}\mathbf{A}_k = 0$, $k > 0$, where $\mathbf{A}_k = 0$ for $k = 0, \dots, P$. Although, instantaneous ICA cannot directly separate convolutive mixtures, there are some cases that an instantaneous ICA can be used to separate convolutive mixture. We will discuss more details in Section 3.

Instantaneous mixing is a special case of convolutive mixing with lag 0. Thus, theoretically, convolutive ICA algorithms are supposed to be able to separate instantaneous mixtures when the sources are temporally independent. However, when the sources are

autocorrelated, it is very challenging to achieve identifiability without appropriate constraints.

In sum, although ICA is based on the minimal independence assumption, algorithms intrinsically assume certain conditions, such as temporal independence. Since such condition may affect the performance of ICA, one has to carefully choose an ICA model among the four possibly scenarios. Ideally, if we can test the temporal independence, the testing results can suggest which ICA model makes sense. However, testing temporal independence is not so trivial. One alternative way to evaluate the existence of autocorrelation is using permutation test on the autocorrelation function (ACF) of $\mathbf{X}$. For each row of $\mathbf{X}$, $X_j(t)$, the

sample autocorrelation function is estimated as $\hat{\rho}_j(h) = \frac{\hat{\gamma}_j(h)}{\hat{\gamma}_j(0)}$, where

$\hat{\gamma}_j(h) = \frac{1}{T} \sum_{t=1}^{T-h} \left(X_j(t+|h|) - \bar{X}_j \right) \left(X_j(t) - \bar{X}_j \right)$. The null distributions of ACF autocorrelation are generated from the permuted $X_j(t)$'s for each $j = 1, \dots, M$, and if any of $\hat{\rho}_j(h)$ is significant, we can say there is an autocorrelation in the system, which implies temporal dependence. Then, using marginal independence based ICA is not encouraged to use. In practical application, testing autocorrelation at each h is not realistic. We suggest to use $h = 1$, although it can be adjusted upon the data.

3. Independent Component Analysis for Echoic Mixing

In this section, we introduce a simple convlutive mixing case wherein the source signals are mixed with different weights over the time, often referred to as “decaying echoic mixing”. Notationally, $\mathbf{A}_k = \mathbf{A}\mathbf{\Theta}_k$, $k = 0, \dots, K$, where $\mathbf{\Theta}_k$ is a diagonal matrix whose diagonal elements represent decaying rates of the sources at time lag k . When the sources are mixed with decaying echoes, we can find a matrix $\mathbf{W}$ such that we can separate the data matrix as a set of independent sources:

$$\begin{aligned} \mathbf{Y}(t) &= (y_1(t), \dots, y_M(t))^T \\ &= \mathbf{W}\mathbf{X}(t) = \sum_{k=0}^P \mathbf{W}\mathbf{A}_k \mathbf{S}(t-k) \\ &= \begin{pmatrix} \theta_{10}s_1(t) + \theta_{11}s_1(t-1) + \dots + \theta_{1p_1}s_1(t-p_1) \\ \vdots \\ \theta_{M0}s_M(t) + \theta_{M1}s_M(t-1) + \dots + \theta_{Mp_M}s_M(t-p_M) \end{pmatrix}. \end{aligned} \quad (3)$$

The instantaneous ICA may separate the mixture marginally, however, the separated sources are autocorrelated. Since s_j 's are white noise sources, (3) can be views as an instantaneous ICA when each source, y_j , is autocorrelated. To solve (3), we employ an existing instantaneous ICA algorithm proposed by [14].

When each source signal is white noise, the separated signals y_j 's (each row of) (3) are moving average (MA) time series with the spectral density function:

$$f_{jj}(r; \theta_j, \sigma_j^2) = \frac{\sigma_j^2}{2\pi} \left| \sum_{k=0}^{pj} \theta_{jk} e^{-ir_k} \right|^2, \quad -\infty < r < \infty.$$

To estimate such $\mathbf{W}$, we consider the Whittle likelihood based method proposed by [14], which utilizes the autocorrelation structures of y_j 's. Suppose that we observe an M vector-valued series $\mathbf{X}(t) = (x_1(t), \dots, x_M(t))^T$, $t = 1, \dots, T$. Each component (row) of $\mathbf{X}$, x_j , $j = 1, \dots, M$, is considered as a univariate time series. Define the *Fourier frequency* by $r_k = \frac{2\pi k}{T}$, $k = 0, \dots, T-1$. Then, the *discrete Fourier transform* (DFT) for the univariate series X_a , $a = 1, \dots, M$, is defined as

$$\varphi(r_k, x_a, T) = \frac{1}{\sqrt{2\pi T}} \sum_{t=1}^T x_a(t) \exp(-ir_k t) = \varphi_a(r_k, \mathbf{X}, T), \quad r_k = \frac{2\pi k}{T}, \quad k=0, \dots, T-1.$$

For the M vector-valued series $\mathbf{X}$, the DFT is defined by

$$\varphi(r_k, \mathbf{X}, T) = \frac{1}{\sqrt{2\pi T}} \sum_{t=1}^T \mathbf{X}(t) \exp(-ir_k t) = \{\varphi_1(r_k, \mathbf{X}, T), \dots, \varphi_M(r_k, \mathbf{X}, T)\}^T,$$

for $r_k = \frac{2\pi k}{T}$, $k = 0, \dots, T-1$.

The *second-order periodogram* of the univariate series x_j is given by

$$\tilde{f}(r_k, x_j, T) = \varphi_j(r_k, \mathbf{X}, T) \bar{\varphi}_j(r_k, \mathbf{X}, T),$$

where $\bar{\varphi}_j$ is the conjugate of a complex valued univariate variable φ_j . For the M vector-valued series $\mathbf{X}$, the second order periodogram is given by

$$\tilde{\mathbf{f}}(r_k, \mathbf{X}, T) = \varphi(r_k, \mathbf{X}, T) \varphi^*(r_k, \mathbf{X}, T)$$

where a vector φ^* is the conjugate transpose of the vector φ .

By asymptotic independence and normality of DFT [27], we can derive the Whittle likelihood in terms of the y_j and their spectral densities:

$$L_T(\theta; \mathbf{Y}) = -\frac{1}{2T} \sum_{j=1}^M \sum_{k=0}^{T-1} \left(\frac{\tilde{f}(r_k, y_j, T)}{f_{jj}(r_k, \theta_j)} + \log f_{jj}(r_k, \theta_j) \right). \quad (4)$$

According to the instantaneous ICA equation and the linearity of DFT, one can derive the Whittle likelihood in terms of the observed mixed signals $\mathbf{X}$ as

$$L_T(\mathbf{W}, \theta; \mathbf{X}) = -\frac{1}{2T} \sum_{j=1}^M \sum_{k=0}^{T-1} \left(\frac{\mathbf{e}_j^T \mathbf{W} \tilde{\mathbf{f}}(\mathbf{r}_k, \mathbf{X}, \mathbf{T}) \mathbf{W}^T \mathbf{e}_j}{f_{jj}(r_k, \theta_j)} + \log f_{jj}(r_k, \theta_j) \right) + \log |\det \mathbf{W}|. \quad (5)$$

The order of MA $\hat{p}_j, j = 1, \dots, M$ are determined using BIC as described in [14]. Once the unmixing matrix and MA parameters $\hat{\Theta}_k = \text{diag} \{ \hat{\theta}_{1k}, \dots, \hat{\theta}_{Mk} \}, k = 0, \dots, P$, where $\hat{\theta}_{jk} = 0, k > \hat{p}_j$, are estimated, we estimate the order of convolutive mixing and convolutive mixing matrices $\{ \hat{\mathbf{A}}_k \}_{k=0, \dots, P}$ as $\hat{P} = \max_{m=1, \dots, M} \hat{p}_m$, and $\hat{\mathbf{A}}_k = \mathbf{W}^{-1} \hat{\Theta}_k$, respectively.

In general, sphering or spatial prewhitening is a commonly used preprocessing technique due to its computational advantages. For the convolutive mixture, the covariance matrix is:

$$\text{Cov}(\mathbf{X}(t)) = \Sigma_X = \sum_{k_1, k_2=0}^P \mathbf{A}_{k_1} \Sigma_S \mathbf{A}_{k_2}^T.$$

After prewhitening, we have $\Sigma_X^{-1/2} \mathbf{X}(t) = \sum_{k=0}^P \Sigma_X^{-1/2} \mathbf{A}_k \mathbf{S}(t-k) = \sum_{k=0}^P \tilde{\mathbf{A}}_k \mathbf{S}(t-k)$.

In this case, $\text{Cov}(\tilde{\mathbf{X}}(t)) = \mathbf{I} = \sum_{k_1, k_2=0}^P \tilde{\mathbf{A}}_{k_1} \Sigma_S \tilde{\mathbf{A}}_{k_2}^T$. With the assumption that $\text{Cov}(\mathbf{S}) = \mathbf{I}$, we have a constraint that $\tilde{\mathbf{A}} = (\tilde{\mathbf{A}}_0, \tilde{\mathbf{A}}_1, \dots, \tilde{\mathbf{A}}_P)$ such that $\tilde{\mathbf{A}} \tilde{\mathbf{A}}^T = \mathbf{I}$. Since $\Theta_k = \mathbf{W} \mathbf{A}_k$, we have $\sum_{k=0}^P \Theta_k^2 = \mathbf{I}$, i.e. for each $j = 1, \dots, M$, the MA coefficient vector has unit norm, $\sum_{k=0}^{p_j} \theta_{jk}^2 = 1$.

After prewhitening with the sample covariance $\tilde{\Sigma}_X = \frac{1}{T} \sum_{t=0}^{T-1} (\mathbf{X}(t) - \bar{\mathbf{X}}) (\mathbf{X}(t) - \bar{\mathbf{X}})^T$,

where the sample mean $\bar{\mathbf{X}} = \frac{1}{T} \sum_{t=0}^{T-1} \mathbf{X}(t)$, we have $\mathbf{Z} = \tilde{\Sigma}_X^{-1/2} \mathbf{X} = \tilde{\Sigma}_X^{-1/2} \mathbf{A} \mathbf{Y} = \mathbf{O} \mathbf{Y}$. With unit variance assumption of $\mathbf{Y}$, the matrix $\mathbf{O}$ is restricted to the Stiefel manifold [28, 29]. We estimate $\mathbf{O}$ via Lagrange multipliers to incorporate the orthogonality constraint.

$$F(\mathbf{O}, \boldsymbol{\lambda}, \theta) = -L_T(\mathbf{O}, \theta; \mathbf{Z}) + \boldsymbol{\lambda}^T \mathbf{C}, \quad (6)$$

where $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_{M(M+1)/2})^T$ is the Lagrange parameter vector, and $\mathbf{C}$ is a $M(M+1)/2$ -dimensional vector with the element being $C_{(j-1)M+k} = (\mathbf{O} \mathbf{O}^T - \mathbf{I}_M)_{jk}, j = 1, \dots, M, k = 1, \dots, j$. Note that we need only $M(M+1)/2$ constraint functions since $\mathbf{O} \mathbf{O}^T - \mathbf{I}_M$ is symmetric. For notational convenience, we write $\boldsymbol{\beta} = (\text{vec}(\mathbf{O}), \boldsymbol{\lambda})$. Write the score and the Hessian matrix as

∇F and $\mathbf{H}$ such that $[\nabla F]_j = \frac{\partial F}{\partial \beta_j}$ and $H_{jk} = \frac{\partial^2 F}{\partial \beta_j \partial \beta_k}$, respectively. The iterative estimation scheme is given below:

The TimeLagICA Algorithm

Pre-whitening $\mathbf{Z} = \hat{\Sigma}_X^{-1/2} \mathbf{X}$.

Initialize $\mathbf{O}^{(old)}, \boldsymbol{\lambda}^{(old)}$.

While the convergence criterion is unsatisfied,

- (1) For given $\mathbf{O}^{(old)}$, estimate the sources: $\hat{\mathbf{Y}} = \mathbf{O}^{(old)} \mathbf{Z}$. For each $j = 1, \dots, M$,
 - (a) Estimate $\hat{\boldsymbol{\theta}}_j$ with constraints $\sum_{k=0}^{p_j} \hat{\theta}_{jk}^2 = 1$.
 - (b) Select the order p_j using BIC.
- (2) For given $\hat{\boldsymbol{\theta}}$, update $\mathbf{O}^{(old)}$ and $\boldsymbol{\lambda}^{(old)}$ as

$$\left(\text{vec}(\mathbf{O}^{(new)}), \boldsymbol{\lambda}^{(new)} \right) = \left(\text{vec}(\mathbf{O}^{(old)}), \boldsymbol{\lambda}^{(old)} \right) - \mathbf{H}^{-1} \nabla F$$

where $\mathbf{H}$ and ∇F depend on $\mathbf{O}^{(old)}$ and $\boldsymbol{\lambda}^{(old)}$.

- (3) $\mathbf{O}^{(old)} \leftarrow \mathbf{O}^{(new)}, \boldsymbol{\lambda}^{(old)} \leftarrow \boldsymbol{\lambda}^{(new)}$

end

Final $\hat{\mathbf{W}} = \mathbf{O} \hat{\Sigma}_X^{-1/2}$. Estimate mixing matrices.

- (1) Estimate the mixing order $\hat{P} = \max_{m=1, \dots, M} p_m$
- (2) Set $\hat{\boldsymbol{\Theta}}_k = \text{diag}\{\hat{\theta}_{1k}, \dots, \hat{\theta}_{Mk}\}, k = 0, \dots, \hat{P}$, where $\hat{\theta}_{jk} = 0, k > p_j$.
- (3) $\hat{\mathbf{A}}_k = \hat{\mathbf{W}}^{-1} \hat{\boldsymbol{\Theta}}_k$

We recover the source $\mathbf{S}$ iteratively using

$$\hat{\mathbf{S}}(t) = \begin{cases} \hat{\mathbf{A}}_0^{-1} \mathbf{X}(t), & t=1 \\ \hat{\mathbf{A}}_0^{-1} \left(\mathbf{X}(t) - \sum_{k=1}^{t-1} \hat{\mathbf{A}}_k \hat{\mathbf{S}}(t-k) \right), & 2 \leq t \leq \hat{P} \\ \hat{\mathbf{A}}_0^{-1} \left(\mathbf{X}(t) - \sum_{k=1}^{\hat{P}} \hat{\mathbf{A}}_k \hat{\mathbf{S}}(t-k) \right), & t > \hat{P}. \end{cases} \quad (7)$$

4. Simulations and Data Analysis

4.1 Simulated Data

In this section, we investigate the effects of model misspecification in ICA. For comparison, we consider three instantaneous ICA algorithms: fastICA [6], JADE [3] and cICA-YW [14] and timelagICA proposed in Section 3. We compare the four ICA algorithms for four different settings: 1) time-lagged convolutive mixture of independent and identically distributed sources (Section 4.1.1), 2) time-lagged convolutive mixing of autoregressive (AR) source (Section 4.1.2), 3) multivariate AR Time Series (Section 4.1.3), 4) instantaneous mixture of temporally autocorrelated sources (Section 4.1.4). As a performance comparison criterion, the correlations between estimated and true sources are used.

4.1.1 Convolutive Mixture of Temporally Independent Sources—We generated 100 datasets with a four dimensional convolutive mixture of four temporally independent and identically distributed sources. The sources were randomly simulated from $N(0, 1)$, centered exponential(1), $t(4)$ and double exponential(1) with sample size $T = 1000$. The sources were mixed as

$$\mathbf{X}(t) = \sum_{k=0}^3 \mathbf{A}_k \mathbf{S}(t-k) = \sum_{k=0}^3 \mathbf{A} \mathbf{\Theta}_k \mathbf{S}(t-k), \quad (8)$$

$$\mathbf{A} = \begin{pmatrix} 0.81 & 0.59 & 0 & 0 \\ -0.59 & 0.81 & 0 & 0 \\ 0 & 0 & 0.72 & 0.70 \\ 0 & 0 & -0.70 & 0.72 \end{pmatrix}, \quad \text{where } \mathbf{\Theta}_1 = \text{diag}\{0.8, 0.5, -0.5, 0.7\}, \mathbf{\Theta}_2 = \text{diag}\{0.64, 0.25, 0, 0\}, \text{ and } \mathbf{\Theta}_3 = \text{diag}\{0.512, 0.125, 0, 0\}.$$

Figure 1 (a) shows the boxplots of the p-values of the temporal dependence test of each observed variable X_j . The red dotted horizontal line represents 0.05. The first two channels show clear temporal dependence at lag 1. This suggests that fastICA and JADE do not work well due to the temporal dependency. For performance comparison, matrices of the absolute correlation between the true sources and estimated sources obtained by four ICA algorithms

are computed. The errors of diagonal elements, $\sqrt{\sum_{j=1}^4 (|Corr(\hat{S}_j, S_j)| - 1)^2}$, and offdiagonal elements, $\sqrt{\sum_{j_1 \neq j_2} (|Corr(\hat{S}_{j_1}, S_{j_2})|)^2}$, are reported in Figure 1 (b) and (c), respectively. The timelagICA performed better than other methods. The off-diagonal elements of the correlation matrices obtained by instantaneous ICA algorithms are small, while the correlation between estimated sources and the true sources are small. The results indicates that instantaneous ICA algorithms cannot specify the time lags although they can separate the convoluted sources marginally.

4.1.2 Convolutional Mixture of Autocorrelated Sources—In this section, we illustrate a case with four dimensional convolutional mixtures of four temporally autocorrelated sources. We generated 100 datasets that the sources were generated from $N(0, 1)$, AR(1) with $\varphi = 0.7$ and standard normal error, ARMA(1,1) with $\varphi = -0.8$, $\theta = 0.5$ and uniform error, and double exponential distribution with mean zero and variance 1 at sample size $T = 1000$. The sources were mixed as (8).

We consider the same algorithms as before. The boxplots of the p-values of the temporal dependence test of each observed variable X_j (Figure 2 (a)) show that all observed variables have significant temporal dependence at lag 1. Figures 1 (b) and (c) show the errors of diagonal and off-diagonal elements of the correlation matrices. The off-diagonal elements of the correlation matrices obtained by timelagICA and cICA-YW are smaller, which indicates that these two methods separate the sources marginally, although they fail to specify time lags. In fact, the traditional instantaneous ICA methods (fastICA and JADE) do not separate first three sources, while they do separate the fourth (double exponential, white noise) source well, which implies these two methods can separate temporally independent sources well although they cannot separate autocorrelated or convoluted mixtures.

4.1.3 Multivariate AR Time Series—In this section we illustrate a case of multivariate AR time series. We generated 100 datasets. In each simulation run, the sources were mixed as:

$$\begin{aligned}
x_1(t) &= S_1(t) \\
x_2(t) &= 0.75x_1(t-1) + S_2(t) \\
x_3(t) &= 0.4x_2(t-1) - 0.75x_1(t-1) + S_3(t),
\end{aligned}$$

where S_1, S_2, S_3 were independently generated from $N(0, 1)$, centered exponential, and double exponential distribution.

Figure 3 (a) shows that the first two variables do not have temporal dependence except two and five simulation runs at 0.05 significance level, respectively. The boxplots of the errors of diagonal elements and offdiagonal elements of the absolute correlation matrices are reported (Figures 3 (b) and (c)) show that none of algorithms separate the sources well under this setting because the ICA model assumptions are violated in the multivariate AR mixing setting.

4.1.4 Instantaneous Mixture of Autocorrelated Sources—In this section we illustrate a case of instantaneous mixing of autocorrelated sources. Four autocorrelated source are generated as follows:

$$\begin{aligned}
S_1(t) &= 0.7S_1(t-1) + \epsilon_1(t), \quad \epsilon_1 \sim U(-\sqrt{3}, \sqrt{3}) \\
S_2(t) &= \epsilon_2(t) + \epsilon_2(t-1) + 0.25\epsilon_2(t-2), \quad \epsilon_2 \sim N(0, 1) \\
S_3(t) &= -0.5S_3(t-1) + \epsilon_3 \sim N(0, 1) \\
S_4(t) &\sim \text{double exponential}(1).
\end{aligned}$$

Figure 4 (a) reports the boxplot of the p-values of the temporal dependence test result of each observation. All variables have significant temporal dependence. The errors of diagonal elements and offdiagonal elements of the absolute correlation matrices (Figures 4 (b) and (c)) show that cICA-YW performed the sources best. Although timelagICA separated sources marginally, it fails to recover time lags due to model misspecification. FastICA and JADE do not separate sources well, particularly the autocorrelated sources generated from normal distribution (the second and the third sources).

4.2 Audio Recording Example

We carry out experiments with 4 music tracks and one simulated noise. We extract 10 seconds in the middle of the tracks and the noise sources are generated from $Noise(t) = 4000 \cos(0.012t + \varphi_1) + 5000 \sin(0.35t + \varphi_2) + 5000Z(t)$, $Z(t) = 0.9Z(t-1) + \epsilon(t)$, $\epsilon \sim N(0, 1)$. All sources are standardized to have zero mean and unit variance. The five sources are displayed in the supplementary document. We mixed the five sources instantaneously using randomly generated mixing matrix $\mathbf{A}$. For the convolutive mixing case, we consider

$$\mathbf{X}(t) = \sum_{k=0}^3 \mathbf{A} \Theta_k \mathbf{S}(t-k),$$

$$\mathbf{A} = \begin{pmatrix} 0.55 & -0.40 & 0.41 & 0.29 & 0.54 \\ -0.22 & 0.74 & 0.37 & 0.45 & 0.25 \\ -0.73 & -0.51 & 0.41 & 0.18 & -0.04 \\ -0.09 & -0.16 & -0.67 & 0.71 & 0.11 \\ -0.33 & 0.03 & -0.28 & -0.42 & 0.80 \end{pmatrix},$$

where $\Theta_1 = \text{diag}\{0.8, -0.5, 0, 0, 0\}$, $\Theta_2 = \text{diag}\{0.64, 0, 0, -0.49, 0\}$, and $\Theta_3 = \text{diag}\{0.512, 0, 0, 0, 0.1\}$.

The temporal autocorrelation test with lag 1 suggests that there are significant temporal dependence for all observations (p-values <0.0001). The absolute correlations between the true and estimated sources are considered as comparison criteria. Figure 5 shows the absolute values of correlations between the estimated sources and the true sources under the instantaneous mixing condition. As we found from the previous simulation studies, cICA-YW performs the best, while other methods do not separate the sources well. The results indicates that the sources have a strong autocorrelation structure, which is not very favorable for the other three ICA algorithms.

The temporal autocorrelation test with lag 1 suggests that there are significant temporal dependence for all observations (p-values <0.0001). Figure 6 shows the absolute values of correlations between true and estimated sources under the convolutive mixing condition. Due to the model ambiguity under the convolutive mixing of autocorrelated sources, the timelagICA and cICA-YW do not separate sources well, while JADE and fastICA performed relatively better. One interesting finding is that the results obtained by fastICA and JADE under both instantaneous and convolutive mixing conditions seem to very similar, while those of timelagICA and cICA-YW are very different. This is mainly due to the intrinsic characteristics of both timelagICA and cICA-YW. Both methods are based on the second-moments of the sources, which may be more dependent on the model specification, while fastICA and JADE are ased on information theory.

4.3 EEG Example

Electroencephalography (EEG) signals measure electrical activity of neurons in the brain. The signals are collected by the electrodes on the scalp. It has been used as a diagnostic tool in many clinical practices. The task we consider this section was largely based on the paradigm from [30] and consisted of the pantomime of 10 common tools (scissors, spoon to stir coffee, ice cream scoop, doorknob, pencil, screwdriver, hammer, paintbrush, key, chalkboard eraser). These tools were selected because the pantomime of their use would allow the participant to keep his or her elbow on the chairs armrest, thus minimizing movement artifact in the EEG recording. Prior to the recording, the participants were asked to demonstrate the correct use of each of the tools. During the EEG recording, the stimuli were presented using evoke software (ANT, the Netherlands). For each trial, subjects first fixated on a cross at the center of the computer monitor; this stimulus lasted four seconds (Figure 2 in the supplementary document). The fixation cross was replaced by the photograph of one of the ten tools with an visual angle of 9 degrees measured from subjects eye to center of the fixation point; participants were instructed not to make any movements during this time. We refer to this first stimulus as the Prepare stimulus. After Prepare remained on the screen for three seconds, a green box appeared around the photograph of the tool (Go stimulus). During the presentation of the Go stimulus (which lasted 3.5

seconds), subjects were to pantomime the use of the tool with their right hand until the word Rest appeared. Rest lasted two seconds and was replaced by the fixation cross. 120 trials were recorded in each subject. Trials with incorrect behavioral performance (negligible number) were marked electronically by a research assistant and were removed from the analysis. The period following the Prepare stimulus was presumed to evoke motor rehearsal, while the period following Go evoked the performance of a tool-use gesture.

The EEG signals were originally collected with 47 electrodes. The dimension was reduced to 6 (99.5% variation explained) via principal component analysis. The temporal autocorrelation tests of the 6 time series were significant at significant level 0.05. We applied four different ICA algorithms to separate sources. Since we do not have ground truth and our goal is to study differences various ICA algorithms, we focus on comparing the four ICA algorithms using absolute correlation clustering.

To see clustering patterns between estimated sources, we applied hierarchical clustering with complete linkage to the absolute value of the correlation matrix. The resulting correlation matrix between the sources recovered by four ICA algorithms is displayed in the supplementary document. The result reveals the ICA algorithms recover similar (i.e. highly correlated) components. As the clustering results suggests, we reorder the components and their topographies. TimelagICA estimates 4 as the order of time lag. The mixing coefficient matrices suggests there is a time delay of order 1. The estimated mixing matrices at different time lags are displayed in the supplementary document.

Figure 7 shows the topography (or spatial pattern) estimated by ICA algorithms. We reordered the topography according to the clustering patterns of the correlations of the estimated sources. For timelagICA, we generated topography of the common mixing matrix A. TimelagICA, cICA-YW and JADE recovered similar spatial components, while fastICA recovered different patterns comparing to other ICA algorithms. The estimated source by the ICA algorithms are displayed in the supplementary document.

5. Discussion

Researchers have had great success in developing efficient ICA algorithms. However, there has not been much research on ICA model validation. This paper is a first step in this direction. In this article, we investigated performance of ICA algorithms under various mixing conditions. Currently existing ICA algorithms are built on assumptions about the nature of the mixing (instantaneous or convolutive) and the nature of the sources (independent or autocorrelated). Our simulation studies show that the performance of ICA algorithms is highly dependent on the mixing conditions and temporal independence of the sources. Most instantaneous ICA algorithms fail to separate autocorrelated sources, while convolutive ICA algorithms seem to highly depend on the model specification and approximation accuracy of unmixing filters.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

The project described was partially supported by grants R01-EB012547 and P41-EB015909 from the National Institute of Biomedical Imaging and Bioengineering, and R01-NS070906 from the National Institute of Neurological Disorders and Stroke.

References

1. Vigário R, Sarela J, Jousmiki V, Hamalainen M, Oja E. Independent component approach to the analysis of eeg and meg recordings. *Biomedical Engineering, IEEE Transactions on*. 2000; 47(5): 589–593.
2. Delorme A, Sejnowski T, Makeig S. Enhanced detection of artifacts in eeg data using higher-order statistics and independent component analysis. *Neuroimage*. 2007; 34(4):1443–1449. [PubMed: 17188898]
3. Cardoso, J.; Souloumiac, A. Radar and signal processing, *iee proceedings f*. Vol. 140. IET; 1993. Blind beamforming for non-gaussian signals; p. 362–370.
4. Bell A, Sejnowski T. An Information-Maximization Approach to Blind Separation and Blind Deconvolution. *Neural Computation*. 1995; 7(6):1129–1159. [PubMed: 7584893]
5. Lee T, Girolami M, Sejnowski T. Independent Component Analysis Using an Extended Infomax Algorithm for Mixed Subgaussian and Supergaussian Sources. *Neural Computation*. 1999; 11(2): 417–441. [PubMed: 9950738]
6. Hyvärinen A, Oja E. Independent Component Analysis: Algorithms and Applications. *Neural Networks*. 2000; 13(4-5):411–30. [PubMed: 10946390]
7. Bach F, Jordan M. Kernel Independent Component Analysis. *Journal of Machine Learning Research*. 2003; 3(1):1–48.
8. Comon P. Independent Component Analysis, a New Concept? *Signal Processing*. 1994; 36(3):287–314.
9. Pham D, Garat P. Blind Separation of Mixture of Independent Sources through a Quasi-maximum Likelihood Approach. *IEEE Transactions on Signal Processing*. 1997; 45(7):1712–1725.
10. Beckmann C, Smith S. Probabilistic Independent Component Analysis for Functional Magnetic Resonance Imaging. *IEEE Transactions on Medical Imaging*. 2004; 23(2):137–152. [PubMed: 14964560]
11. Palmer, J.; Makeig, S.; Delgado, K.; Rao, B. Newton Method for the ICA Mixture Model; Acoustics, Speech and Signal Processing, 2008 ICASSP 2008 IEEE International Conference on; 2008. p. 1805–1808.
12. Chen A, Bickel P. Consistent Independent Component Analysis and Prewhitening. *IEEE Transactions on Signal Processing*. 2005; 53(10 Part 1):3625–3632.
13. Matteson, D.; Tsay, R. Independent component analysis via distance covariance. Cornell University; 2011. Tech Rep
14. Lee S, Shen H, Truong Y, Lewis M, Huang X. Independent Component Analysis Involving Autocorrelated Sources with an Application to functional Magnetic Resonance Imaging. *Journal of the American Statistical Association*. 2011; 106:1009–1024.
15. Anemüller J, Kollmeier B. Adaptive separation of acoustic sources for anechoic conditions: A constrained frequency domain approach. *Speech Communication*. 2003; 39(1-2):79–95.
16. Bofill P. Underdetermined blind separation of delayed sound sources in the frequency domain. *Neurocomputing*. 2003; 55(3):627–641.
17. Wu, H.; Principe, J. Simultaneous diagonalization in the frequency domain (sdif) for source separation; *Proc. ica*. Citeseer; 1999. p. 245–250.
18. Chen B, Petropulu A. Frequency domain blind mimo system identification based on second and higher order statistics. *Signal Processing, IEEE Transactions on*. 2001; 49(8):1677–1688.
19. Diamantaras K, Petropulu A, Chen B. Blind two-input-two-output fir channel identification based on frequency domain second-order statistics. *Signal Processing, IEEE Transactions on*. 2000; 48(2):534–542.

20. Castella M, Pesquet J, Petropulu A. A family of frequency-and time-domain contrasts for blind separation of convolutive mixtures of temporally dependent signals. *Signal Processing, IEEE Transactions on*. 2005; 53(1):107–120.
21. Gorokhov, A.; Loubaton, P.; Moulines, E. *Acoustics, speech, and signal processing, 1996. icassp-96. conference proceedings., 1996 ieee international conference on*. Vol. 5. IEEE; 1996. Second order blind equalization in multiple input multiple output fir systems: A weighted least squares approach; p. 2415-2418.
22. Icart, S.; Gautier, R. *Acoustics, speech, and signal processing, 1996. icassp-96. conference proceedings., 1996 ieee international conference on*. Vol. 5. IEEE; 1996. Blind separation of convolutive mixtures using second and fourth order moments; p. 3018-3021.
23. Pearlmutter B, Parra L. Maximum Likelihood Source Separation: a Context-sensitive Generalization of ICA. *Advances in Neural Information Processing Systems*. 1997; 9:613–619.
24. Dyrholm M, Makeig S, Hansen L. Model Selection for Convolutive ICA with an Application to Spatiotemporal Analysis of EEG. *Neural Computation*. 2007; 19(4):934–955. [PubMed: 17348768]
25. O'Grady P, Pearlmutter B, Rickard S. Survey of sparse and non-sparse methods in source separation. *International Journal of Imaging Systems and Technology*. 2005; 15(1):18–33.
26. Comon, P.; Jutten, C. *Handbook of Blind Source Separation: Independent Component Analysis and Applications*. Academic Press; 2010.
27. Brillinger, D. *Society for Industrial and Applied Mathematics*. 2nd. 2001. *Time Series: Data Analysis and Theory*.
28. Edelman A, Arias T, Smith S. The Geometry of Algorithms with Orthogonality Constraints. *SIAM Journal on Matrix Analysis and Applications*. 1998; 20:303–353.
29. Amari S. Natural Gradient Learning for Over-and Under-complete Bases in ICA. *Neural Computation*. 1999; 11(8):1875–1883. [PubMed: 10578035]
30. Wheaton L, Shibasaki H, Hallett M. Temporal activation pattern of parietal and premotor areas related to praxis movements. *Clinical neurophysiology*. 2005; 116(5):1201–1212. [PubMed: 15826863]

Figure 1.

Convulsive Mixture of Temporally Independent Sources: The temporal dependence test result of each observed channel X_j is reported in (a). The first two channels show clear temporal dependence with lag 1. Matrices of the absolute correlation between the true sources and estimated sources obtained by four ICA algorithms. The errors of diagonal elements and offdiagonal elements are reported in (b) and (c), respectively. The timelagICA performs better than other methods.

Figure 2.

Convulsive Mixture of Temporally Independent Sources: The temporal dependence test result of each observed channel X_j is reported in (a). All variables had significant temporal dependence with lag 1. Matrices of the absolute correlation between the true sources and estimated sources by four ICA algorithms were computed. The errors of diagonal elements and offdiagonal elements are reported in (b) and (c), respectively.

Figure 3.

Multivariate AR mixture: The temporal dependence test result of each observed channel X_j is reported in (a). First two variables do not have temporal dependence except two and five simulation runs at 0.05 significance level, respectively. The errors of diagonal elements and offdiagonal elements of the absolute correlation matrices are reported in (b) and (c), respectively.

Figure 4.

Instantaneous Mixture of Autocorrelated Sources: The temporal dependence test result of each observed channel X_j is reported in (a). All variables have significant temporal dependence. The errors of diagonal elements and off-diagonal elements of the absolute correlation matrices are reported in (b) and (c), respectively. The cICA-YW performed the best.

Figure 5.

Instantaneous Mixtures: As we found from the previous simulation studies, cICA-YW performs the best, while other methods do not separate the sources well.

Figure 6.

Convolved Mixtures: timelagICA and cICA-YW do not separate sources well, while JADE and fastICA performed relatively better.

Figure 7.
Spatial maps. We reordered the topography according to the clustering patterns of the correlations of the estimated sources.