



A fixed-point algorithm for blind source separation with nonlinear autocorrelation[☆]

Zhenwei Shi^{*}, Zhiguo Jiang, Fugen Zhou

Image Processing Center, School of Astronautics, Beijing University of Aeronautics and Astronautics, Beijing 100083, PR China

ARTICLE INFO

Article history:

Received 8 December 2007

Received in revised form 8 March 2008

Keywords:

Blind source separation (BSS)

Independent component analysis (ICA)

Nonlinear autocorrelation

Fixed-point algorithm

ABSTRACT

This paper addresses blind source separation (BSS) problem when source signals have the temporal structure with nonlinear autocorrelation. Using the temporal characteristics of sources, we develop an objective function based on the nonlinear autocorrelation of sources. Maximizing the objective function, we propose a fixed-point source separation algorithm. Furthermore, we give some mathematical properties of the algorithm. Computer simulations for sources with square temporal autocorrelation and the real-world applications in the analysis of the magnetoencephalographic recordings (MEG) illustrate the efficiency of the proposed approach. Thus, the presented BSS algorithm, which is based on the nonlinear measure of temporal autocorrelation, provides a novel statistical property to perform BSS.

© 2008 Elsevier B.V. All rights reserved.

1. Introduction

Recently, blind source separation (BSS) [6,11], as an increasingly popular data analysis technique, has received wide attention in various fields such as biomedical signal processing and analysis, speech and image processing, wireless telecommunication systems, data mining, etc. The main task of BSS is to recover original sources from their mixtures using some statistical properties of original sources.

The blind source separation problem has been studied by researchers in applied mathematics, neural networks and statistical signal processing. Several methods for BSS using the statistical properties of original sources have been proposed, such as non-Gaussianity (or equivalently, independent component analysis, ICA) [1,3,5–8,11–13,24,34], linear predictability or smoothness [2,6], linear autocorrelation [4,19,31], coding complexity [9,26,27,29], temporal predictability [23], nonstationarity [10,18,21], sparsity [14,15,25,35], energy predictability [28], nonlinear innovation [30] and nonnegativity [20,22], etc.

In this paper, we present a way using nonlinear autocorrelation of source signals for BSS. We propose a fixed-point algorithm for BSS based on nonlinear autocorrelation and analyze its stability conditions. We show that the BSS problem can be solved by maximizing nonlinear temporal autocorrelation of the sources. When sources have square temporal autocorrelation, we demonstrate that the efficient implementation of the method. Also, the proposed algorithm can extract the most interesting and meaningful sources in brain magnetoencephalography (MEG) data analysis. Thus, the presented BSS algorithm, which is based on the nonlinear measure of temporal autocorrelation, provides a new statistical property to perform BSS.

The structure of the paper is as follows. The objective function based on the nonlinear autocorrelation of the desired source signals, and a fixed-point algorithm for optimizing the objective function are proposed in Section 2. Furthermore, we analyze the stability conditions of the fixed-point algorithm in Section 3. In Section 4, experiments on different datasets are presented. Some conclusions are drawn in Section 5.

[☆] The work was supported by the National Science Foundation of China (60605002, 60776793, 10571018).

^{*} Corresponding author. Tel.: +86 10 627 96 872; fax: +86 10 627 86 911.

E-mail address: shizhenwei@mail.tsinghua.edu.cn (Z. Shi).

2. Proposed algorithm

2.1. Objective function

We observe sensor signals $\mathbf{x}(t) = (x_1(t), \dots, x_n(t))^T$ described in the matrix notation:

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

where \mathbf{A} is an $n \times n$ unknown mixing matrix, $\mathbf{s}(t) = (s_1(t), \dots, s_n(t))^T$ is a vector of unknown zero-mean and unit-variance primary sources.

The basic problem of BSS is then to estimate both the mixing matrix \mathbf{A} and the source signals $s_i(t)$ using only observations of the mixtures $x_i(t)$ ($i = 1, \dots, n$), i.e., to find an $n \times n$ separating matrix $\mathbf{W} = (\mathbf{w}_1, \dots, \mathbf{w}_n)^T$ such that the unmixed signals $\mathbf{y}(t) = (y_1(t), \dots, y_N(t))^T$,

$$\mathbf{y}(t) = \mathbf{W}\mathbf{x}(t) \tag{2}$$

are the estimated source signals. The sources are recovered up to scaling and permutation.

A useful preprocessing strategy in BSS is to first whiten the observed variables [11]. This means that we transform the observed vector \mathbf{x} linearly so that we obtain a new vector $\tilde{\mathbf{x}}$ which is white, i.e., its components are uncorrelated and their variances equal unity. In other words, \mathbf{x} is linearly transformed into a vector

$$\tilde{\mathbf{x}}(t) = \mathbf{V}\mathbf{x}(t) = \mathbf{V}\mathbf{A}\mathbf{s}(t) \tag{3}$$

whose covariance matrix equals the identity matrix: $E\{\tilde{\mathbf{x}}(t)\tilde{\mathbf{x}}(t)^T\} = \mathbf{I}$, where \mathbf{V} is a whitening matrix.

If we want to estimate a desired source signal, for this purpose we design a single processing unit described as

$$\tilde{y}_i(t) = \mathbf{w}_i^T \tilde{\mathbf{x}}(t), \tag{4}$$

$$\tilde{y}_i(t - \tau_k) = \mathbf{w}_i^T \tilde{\mathbf{x}}(t - \tau_k), \tag{5}$$

where $\mathbf{w}_i = (w_{i1}, \dots, w_{in})^T$ is the weight vector which corresponds to the estimate of one row of $(\mathbf{V}\mathbf{A})^{-1}$, and $\tilde{y}_i(t)$ is the output signal which corresponds to the estimate of the source signal $s_i(t)$, and τ_k is a certain time delay.

We present the following constrained maximization problem based on several time delay nonlinear autocorrelation function of the desired source:

$$\begin{aligned} \max_{\|\mathbf{w}_i\|=1} \psi_1(\mathbf{w}_i) &= \sum_{k=1}^M E\{G(\tilde{y}_i(t))G(\tilde{y}_i(t - \tau_k))\} \\ &= \sum_{k=1}^M E\{G(\mathbf{w}_i^T \tilde{\mathbf{x}}(t))G(\mathbf{w}_i^T \tilde{\mathbf{x}}(t - \tau_k))\}. \end{aligned} \tag{6}$$

G is a differentiable nonlinear function, which measures the nonlinear autocorrelation degree of the desired source. Examples of choices are $G(u) = u^2$ or $G(u) = \log \cosh(u)$.

Sometimes, we could use only one time lagged nonlinear autocorrelation to obtain the desired source signal:

$$\begin{aligned} \max_{\|\mathbf{w}_i\|=1} \psi_2(\mathbf{w}_i) &= E\{G(\tilde{y}_i(t))G(\tilde{y}_i(t - \tau_k))\} \\ &= E\{G(\mathbf{w}_i^T \tilde{\mathbf{x}}(t))G(\mathbf{w}_i^T \tilde{\mathbf{x}}(t - \tau_k))\}. \end{aligned} \tag{7}$$

Thus, the constrained maximization problem (7) is a special case of the problem (6) when only one time delay can be used. In Section 3, we give some theoretical properties about the maximization problems (6) and (7).

2.2. Learning algorithms

Maximizing the objective function in (6), we derive a fixed-point blind source separation (BSS) algorithm. The gradient of $\psi_1(\mathbf{w}_i)$ with respect to \mathbf{w}_i can be obtained as

$$\frac{\partial \psi_1(\mathbf{w}_i)}{\partial \mathbf{w}_i} = \sum_{k=1}^M E\{g(\tilde{y}_i(t))G(\tilde{y}_i(t - \tau_k))\tilde{\mathbf{x}}(t) + G(\tilde{y}_i(t))g(\tilde{y}_i(t - \tau_k))\tilde{\mathbf{x}}(t - \tau_k)\}. \tag{8}$$

The function g is the derivative of G . According to the Kuhn–Tucker conditions [17], we note that at a stable point of the optimization problem (6), the gradient of $\psi_1(\mathbf{w}_i)$ at \mathbf{w}_i must point in the direction of \mathbf{w}_i , thus we can optimize the objective function in (6) by a fixed-point algorithm [11]. This means that we have

$$\begin{aligned} \mathbf{w}_i &\leftarrow \sum_{k=1}^M E\{g(\tilde{y}_i(t))G(\tilde{y}_i(t - \tau_k))\tilde{\mathbf{x}}(t) + G(\tilde{y}_i(t))g(\tilde{y}_i(t - \tau_k))\tilde{\mathbf{x}}(t - \tau_k)\}, \\ \mathbf{w}_i &\leftarrow \mathbf{w}_i / \|\mathbf{w}_i\|. \end{aligned} \tag{9}$$

If only one time delay can be used, i.e. maximizing the objective function in (7), we have

$$\begin{aligned} \mathbf{w}_i &\leftarrow E\{g(\tilde{y}_i(t))G(\tilde{y}_i(t - \tau_k))\tilde{\mathbf{x}}(t) + G(\tilde{y}_i(t))g(\tilde{y}_i(t - \tau_k))\tilde{\mathbf{x}}(t - \tau_k)\}, \\ \mathbf{w}_i &\leftarrow \mathbf{w}_i / \|\mathbf{w}_i\|. \end{aligned} \tag{10}$$

To estimate several source signals, one can simply use a deflation scheme (Gram-Schmidt orthogonalization scheme) or the symmetric orthogonalization [11].

Thus, the fixed-point BSS algorithm (**FixNA**- BSS using the **Fixed**-point algorithm for maximizing the **Nonlinear Autocorrelation**) is obtained as follows:

Algorithm outline: FixNA (estimating one source)

- (1) Center the data to make its mean zero and whiten the data to give $\tilde{\mathbf{x}}(t)$. Choose random initial values for \mathbf{w}_i , τ_k , and M .
- (2) Update the weight vector by

$$\begin{aligned} \mathbf{w}_i &\leftarrow \sum_{k=1}^M E\{g(\tilde{y}_i(t))G(\tilde{y}_i(t - \tau_k))\tilde{\mathbf{x}}(t) + G(\tilde{y}_i(t))g(\tilde{y}_i(t - \tau_k))\tilde{\mathbf{x}}(t - \tau_k)\}, \\ \mathbf{w}_i &\leftarrow \mathbf{w}_i / \|\mathbf{w}_i\|. \end{aligned} \tag{11}$$

- (3) If not converged, go back to step (2).

Note that convergence of the fixed-point algorithm means that the old and new values of \mathbf{w}_i point in the same direction, as the well-know fixed-point ICA algorithms [8,11].

To estimate the separating matrix $\mathbf{W} = (\mathbf{w}_1, \dots, \mathbf{w}_n)^T$ or estimate all the source signals, one can simply use a deflation scheme (one-by-one estimation) or the symmetric orthogonalization [11].

3. Theoretical properties of the FixNA

We analyze the stability conditions about the FixNA in this section.¹

Theorem 1. Assume that the input data follows the model (1) with whitened data $\tilde{\mathbf{x}} = \mathbf{V}\mathbf{s}$ where \mathbf{V} is the whitening matrix, and $\Psi_1(\mathbf{w})$ is a sufficiently smooth even function. Furthermore, assume that $\{s_i, s_{i, \tau_k}\}$ and $\{s_j, s_{j, \tau_k}\}$ ($\forall j \neq i$) are mutually independent. Then the local maxima of $\Psi_1(\mathbf{w})$ under the constraint $\|\mathbf{w}\| = 1$ include one row of the inverse of the mixing matrix $\mathbf{V}\mathbf{A}$ such that the corresponding desired source signal s_i satisfy

$$\begin{aligned} \sum_{k=1}^M E\{g'(s_i)G(s_{i, \tau_k}) + g'(s_{i, \tau_k})G(s_i) + 2s_j s_{j, \tau_k} g(s_i)g(s_{i, \tau_k}) \\ - s_i g(s_i)G(s_{i, \tau_k}) - s_{i\tau} g(s_{i, \tau_k})G(s_i)\} < 0, \quad (\forall j \neq i), \end{aligned} \tag{12}$$

where g is the derivative of G , and g' is the derivative of g .

Proof. Assume that $\{s_i, s_{i, \tau_k}\}$ and $\{s_j, s_{j, \tau_k}\}$ ($\forall j \neq i$) are mutually independent, and the vector $\tilde{\mathbf{x}}$ is white, we have:

$$E\{\tilde{\mathbf{x}}\tilde{\mathbf{x}}^T\} = \mathbf{V}\mathbf{A}E\{\mathbf{s}\mathbf{s}^T\}\mathbf{A}^T\mathbf{V}^T = (\mathbf{V}\mathbf{A})(\mathbf{V}\mathbf{A})^T = \mathbf{I}, \tag{13}$$

which means the matrix $\mathbf{V}\mathbf{A}$ is orthogonal. Make the change of coordinates $\mathbf{p} = (p_1, \dots, p_n)^T = \mathbf{A}^T\mathbf{V}^T\mathbf{w}$ and note that $\mathbf{w}^T\tilde{\mathbf{x}} = \mathbf{w}^T\mathbf{V}\mathbf{A}\mathbf{s}$, we have:

$$\Psi_1(\mathbf{p}) = \sum_{k=1}^M E\{G(\mathbf{p}^T\mathbf{s})G(\mathbf{p}^T\mathbf{s}_{\tau_k})\}. \tag{14}$$

It is enough to analyze the stability of the point $\mathbf{p} = \mathbf{e}_1$, where $\mathbf{e}_1 = (1, 0, 0, \dots)^T$ (because $\Psi_1(\mathbf{w})$ is even, nothing changes for $(-1, 0, 0, \dots)^T$). Evaluating the gradient and the Hessian at point $\mathbf{p} = \mathbf{e}_1$ and using the independency assumptions, and making a small perturbation $\varepsilon = (\varepsilon_1, \varepsilon_2, \dots)^T$, we obtain:

$$\begin{aligned} \Psi_1(\mathbf{e}_1 + \varepsilon) &= \Psi_1(\mathbf{e}_1) + \varepsilon^T \frac{\partial \Psi_1(\mathbf{e}_1)}{\partial \mathbf{p}} + \frac{1}{2} \varepsilon^T \frac{\partial^2 \Psi_1(\mathbf{e}_1)}{\partial \mathbf{p}^2} \varepsilon + o(\|\varepsilon\|^2) \\ &= \Psi_1(\mathbf{e}_1) + \sum_{k=1}^M E\{s_1 g(s_1)G(s_{1, \tau_k}) + s_{1\tau} g(s_{1, \tau_k})G(s_1)\} \varepsilon_1 \\ &\quad + \sum_{k=1}^M \frac{1}{2} [E\{s_1^2 g'(s_1)G(s_{1, \tau_k}) + s_{1, \tau_k}^2 g'(s_{1, \tau_k})G(s_1) + 2s_1 s_{1, \tau_k} g(s_1)g(s_{1, \tau_k})\}] \varepsilon_1^2 \\ &\quad + \frac{1}{2} \sum_{k=1}^M \sum_{j>1} [E\{g'(s_1)G(s_{1, \tau_k}) + g'(s_{1, \tau_k})G(s_1) + 2s_j s_{j, \tau_k} g(s_1)g(s_{1, \tau_k})\}] \varepsilon_j^2 + o(\|\varepsilon\|^2). \end{aligned} \tag{15}$$

¹ We will drop the sampling index t for simplicity, i.e., $y_{\tau_k} = y(t - \tau_k)$, $s_{i, \tau_k} = s_i(t - \tau_k)$, and $\mathbf{w} = \mathbf{w}_i$.

Due to the constraint $\|\mathbf{w}\| = 1$ and the matrix \mathbf{VA} is orthogonal, we have $\|\mathbf{p}\| = \|\mathbf{A}^T \mathbf{V}^T \mathbf{w}\| = 1$. Thus, we get $\varepsilon_1 = \sqrt{1 - \varepsilon_2^2 - \varepsilon_3^2 - \dots} - 1$. Due to the fact that $\sqrt{1 - \alpha} = 1 - \frac{\alpha}{2} + o(\alpha)$, the term of order ε_1^2 is $o(\|\varepsilon\|^2)$, i.e., of higher order, and can be neglected [11]. Using the first-order approximation for ε_1 , we obtain $\varepsilon_1 = -\sum_{j>1} \frac{\varepsilon_j^2}{2} + o(\|\varepsilon\|^2)$ [11], which finally gives

$$\begin{aligned} \psi_1(\mathbf{e}_1 + \varepsilon) &= \psi_1(\mathbf{e}_1) + \frac{1}{2} \sum_{k=1}^M \sum_{j>1} E\{g'(s_1)G(s_{1,\tau_k}) + g'(s_{1,\tau_k})G(s_1) \\ &\quad + 2s_j s_{j,\tau_k} g(s_1)g(s_{1,\tau_k}) - s_1 g(s_1)G(s_{1,\tau_k}) - s_{1,\tau_k} g(s_{1,\tau_k})G(s_1)\} \varepsilon_j^2 + o(\|\varepsilon\|^2) \end{aligned} \tag{16}$$

which clearly proves $\mathbf{p} = \mathbf{e}_1$ is an extremum, and of the type implied by the condition of the theorem. \square

If we choose only one time delay, we can similarly prove the following corollary.

Corollary 1. Assume that the input data follows the model (1) with whitened data $\tilde{\mathbf{x}} = \mathbf{VA}\mathbf{s}$ where \mathbf{V} is the whitening matrix, and $\psi_2(\mathbf{w})$ is a sufficiently smooth even function. Furthermore, assume that $\{s_i, s_{i,\tau_k}\}$ and $\{s_j, s_{j,\tau_k}\}$ ($\forall j \neq i$) are mutually independent. Then the local maxima of $\psi_2(\mathbf{w})$ under the constraint $\|\mathbf{w}\| = 1$ include one row of the inverse of the mixing matrix \mathbf{VA} such that the corresponding desired source signal s_i satisfy

$$E\{g'(s_i)G(s_{i,\tau_k}) + g'(s_{i,\tau_k})G(s_i) + 2s_j s_{j,\tau_k} g(s_i)g(s_{i,\tau_k}) - s_i g(s_i)G(s_{i,\tau_k}) - s_{i,\tau_k} g(s_{i,\tau_k})G(s_i)\} < 0, \quad (\forall j \neq i), \tag{17}$$

where g is the derivative of G , and g' is the derivative of g .

Remark 1. Assume that the following conditions are satisfied ($\forall j \neq i$): (a) $\{s_i, s_{i,\tau_k}\}$ and $\{s_j, s_{j,\tau_k}\}$ are mutually independent; (b) $E\{s_j s_{j,\tau_k}\} = 0$, from (12), we have the further stability condition:

$$\sum_{k=1}^M E\{g'(s_i)G(s_{i,\tau_k}) + g'(s_{i,\tau_k})G(s_i) - s_i g(s_i)G(s_{i,\tau_k}) - s_{i,\tau_k} g(s_{i,\tau_k})G(s_i)\} < 0, \quad (\forall j \neq i). \tag{18}$$

Particularly, assume that the square function $G(u) = u^2$ is chosen, we have the stability condition $\sum_{k=1}^M \text{cov}(s_i^2, s_{i,\tau_k}^2) > 0$.

Remark 2. Assume that the signal has no time dependencies and $G(s_{i,\tau_k}) = 1$, from the stability condition (17), we obtain:

$$E\{g'(s_i) - s_i g(s_i)\} < 0. \tag{19}$$

This is in fact the well-known stability condition of independent component analysis (ICA) [11].

4. Experimental results

4.1. Experimental results about source signals with square temporal autocorrelations

We demonstrate that when we choose $G(u) = u^2$ and $G(u) = \log(\cosh(u))$, the two FixNA algorithms² can separate the source signals with square temporal autocorrelations.

We create five artificial signals which have square temporal autocorrelations as follows (with Gaussian marginal distributions, zero linear autocorrelations)³ [10]: first, we create five signals using a first-order autoregressive model with constant variances of the innovations, with 5000 time points. The signals are created with Gaussian innovations and have identical autoregressive coefficients (0.8). All these innovations have constant unit variance. Then, the signs of the signals are completely randomized by multiplying each signal by a binary i.i.d. signal that takes the values ± 1 with equal probabilities. The five source signals are shown in Fig. 1. The source signals are mixed with 5×5 random matrices. The FixNA algorithms with the nonlinear functions ($G(u) = u^2$ and $G(u) = \log(\cosh(u))$) with the symmetric orthogonalization are used to estimate the separating matrix. For comparison, we also run the cumulant-based fixed-point approach using the nonstationarity of variance (FPNSV) ($\tau_k = 1$) [10]. In order to measure the accuracy of separation, we calculate the performance index [1,16,28,30]

$$\begin{aligned} \text{PI} &= \frac{1}{n^2} \left\{ \sum_{i=1}^n r\text{PI}_i + \sum_{j=1}^n c\text{PI}_j \right\} \\ &= \frac{1}{n^2} \left\{ \sum_{i=1}^n \left(\sum_{j=1}^n \frac{|p_{ij}|}{\max_k |p_{ik}|} - 1 \right) + \sum_{j=1}^n \left(\sum_{i=1}^n \frac{|p_{ij}|}{\max_k |p_{kj}|} - 1 \right) \right\}, \end{aligned} \tag{20}$$

² We choose the initial values $\tau_k = 1$ and $M = 1$ for the two FixNA algorithms, which are to be suitable for the all experiments.

³ Thus, source signals of this kind could not be separated by ordinary source separation methods based on non-Gaussianity, such as FastICA [11], linear correlation method such as AMUSE [31] and SOBI [4].

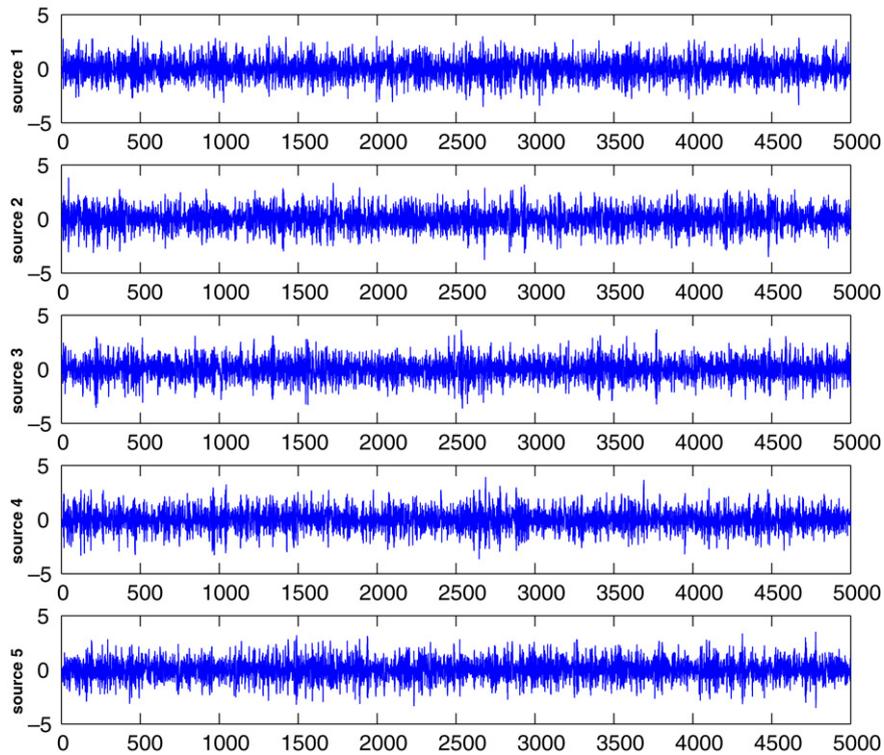


Fig. 1. Five sources with square temporal autocorrelation.

where $r\mathbf{PI}_i = \sum_{j=1}^n \frac{|p_{ij}|}{\max_k |p_{ik}|} - 1$ and $c\mathbf{PI}_j = \sum_{i=1}^n \frac{|p_{ij}|}{\max_k |p_{kj}|} - 1$ in which p_{ij} is the ij th element of $n \times n$ matrix $\mathbf{P} = \mathbf{WVA}$. The term $r\mathbf{PI}_i$ gives the error of the separation of the output component y_i with respect to the sources and $c\mathbf{PI}_j$ measures the degree of the source c_j appearing multiple times at the output. The larger the value \mathbf{PI} is, the poorer the statistical performance of a BSS algorithm [1,16,28,30].

Fig. 2 shows the average performance indexes over 100 independent trials against iteration numbers by the three algorithms. The FixNA algorithm using $G(u) = u^2$ performs similarly to the FPNSV algorithm in the sense of the separation accuracy. The FastNA algorithm using $G(u) = \log(\cosh(u))$ has a slightly better performance than the other two algorithms.

To investigate the robustness of algorithms, we randomly add 30 outliers whose values are 10 in each source signal. Fig. 3 shows the average performance indexes over 100 independent trials against iteration numbers. Obviously, the FixNA algorithm using $G(u) = \log(\cosh(u))$ outperforms the two other algorithms when the outliers are introduced. In fact, the FPNSV algorithm and the FixNA algorithm using $G(u) = u^2$ are both sensitive to the outliers and degrade in performance. Thus, the FixNA algorithm with the nonlinear $G(u) = \log(\cosh(u))$ could have better convergence in practice.

4.2. Applications in magnetoencephalography (MEG) data analysis

Magnetoencephalography (MEG) is a passive functional brain imaging technique which, under ideal conditions, can monitor the activation of a neuronal population with a spatial resolution of a few mm and with millisecond temporal resolution. When performing MEG measurements, physicians often have to deal with considerable amounts of artifacts that may render impossible the extraction of valuable information therein. The presence of artifacts, such as eye and muscle activity, the activity of heart and environment electric, digital watch and magnetic disturbances, may render impossible the extraction of useful information from MEG data. If the magnitude of artifacts is larger than that of brain signals, it may become hard to distinguish and interpret the actual and valuable brain activity [32,33]. It is challenging to identify and remove the artifacts from the signals and analyze/explain the brain signals themselves.

Independent component analysis (ICA) has been shown effective to extract and eliminate the artifacts [32,33]. Unlike ICA which assumes that the source signals are non-Gaussian, we use the nonlinear autocorrelation to capture the statistical character of sources. Several artifact structures are evident, such as eye and muscle activity. Thus, the nonlinear temporal autocorrelation is also likely a strong feature of MEG data sets.

The MEG data set studied in this section consists of preprocessed signals originating from 122-channel whole-scalp MEG measurements from the brain [32,33]. The original signals are band-pass filtered between 0.5 and 45 Hz, and the data dimension is reduced from 122 to 20 using the principal component analysis (PCA) whitening approach to reduce noise and over-learning. The record lasts 2 min and includes 17730 samples and the 20 whitened signals are illustrated in Fig. 4.

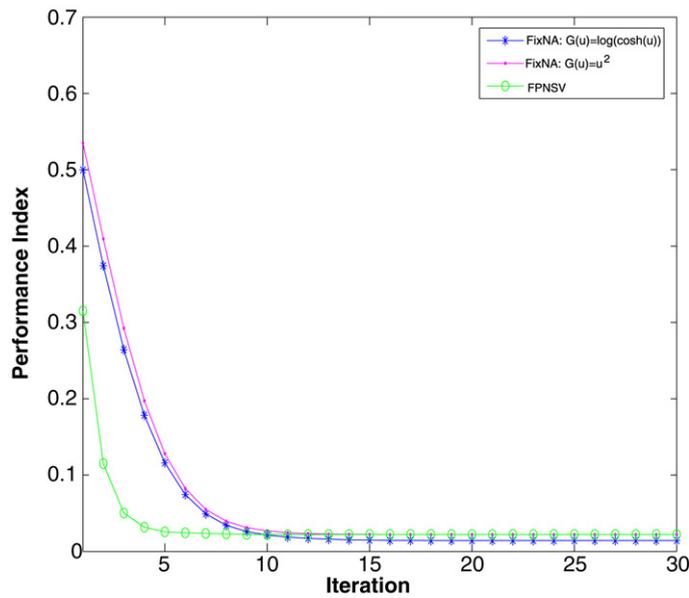


Fig. 2. Average performance indexes over 100 independent runs for 5 sources with square temporal autocorrelation.

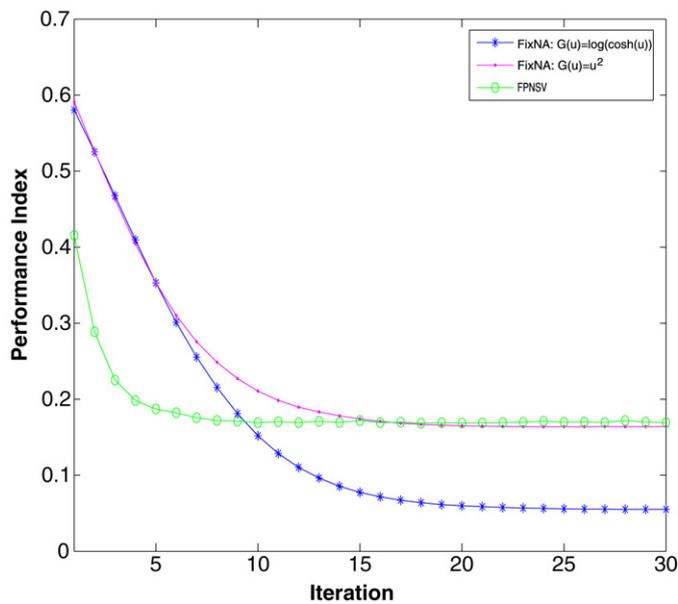


Fig. 3. Average performance indexes over 100 independent runs for 5 sources with square temporal autocorrelation when the outliers are introduced.

Fig. 5 shows the extracted 20 nonlinear autocorrelation source signals with the proposed FixNA approach (using $G(u) = \log(\cosh(u))$). It is known that some of extracted source signals have clear physical or physiological interpretation [32,33], i.e. the extracted source signals 5 and 16 correspond to eye movements, 7 and 8 are related to the muscular activities, the signal 2 corresponds to the heart beating, and 17 shows clearly the artifact originated from the digital watch.

5. Conclusions

We have proposed the FixNA algorithm for blind source separation based on the nonlinear autocorrelation of source signals and have analyzed its stability conditions. We show that the BSS problem can be solved by maximizing the nonlinear temporal autocorrelation of sources. When sources have square temporal autocorrelation, we demonstrated the efficient implementation of the method. Also, the proposed FixNA can extract the artifacts in brain MEG data analysis. Thus, the FixNA algorithm, which is based on the nonlinear measure of temporal autocorrelations of sources, provides a novel statistical property to perform BSS.

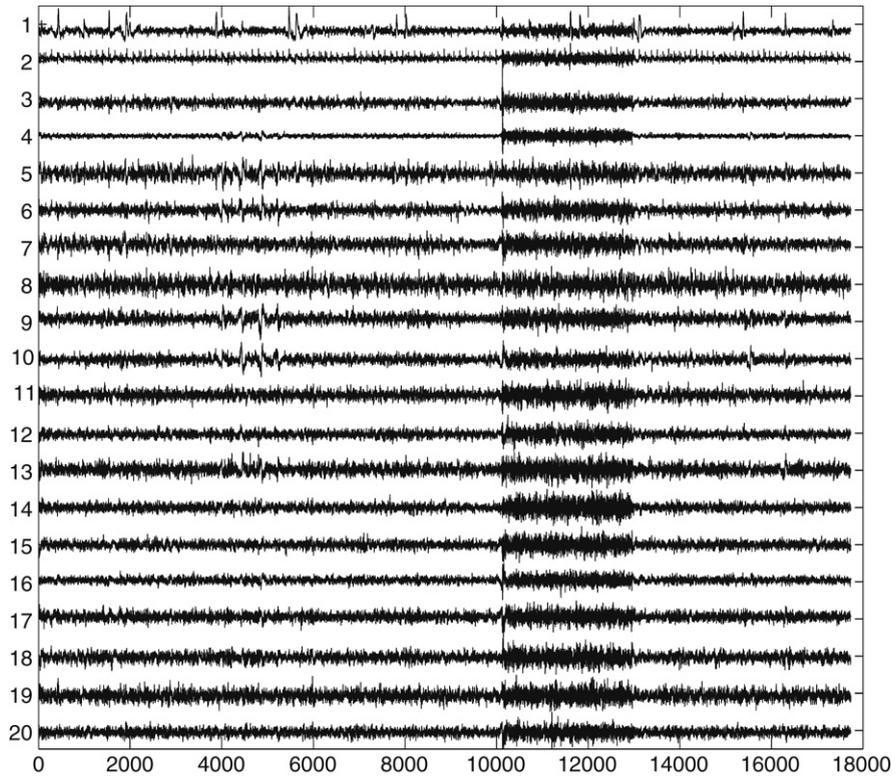


Fig. 4. The 20 whitened MEG signals.

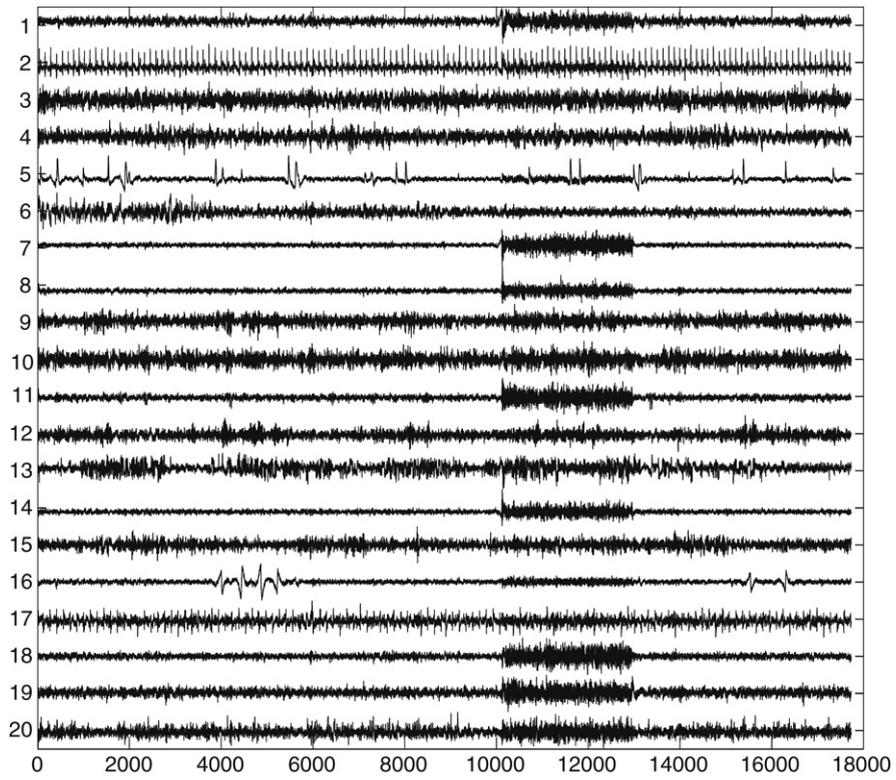


Fig. 5. The 20 source signals obtained by the presented FixNA.

Acknowledgments

The authors wish to gratefully thank all anonymous reviewers who provided insightful and helpful comments.

References

- [1] S.I. Amari, A. Cichocki, H. Yang, A new learning algorithm for blind source separation, in: *Advances in Neural Information Processing Systems*, vol. 8, 1996, pp. 757–763.
- [2] A.K. Barros, A. Cichocki, Extraction of specific signals with temporal structure, *Neural Computation* 13 (9) (2001) 1995–2003.
- [3] A. Bell, T. Sejnowski, An information-maximization approach to blind separation and blind deconvolution, *Neural Computation* 7 (6) (1995) 1129–1159.
- [4] A. Belouchrani, K.A. Meraim, J.-F. Cardoso, E. Moulines, A blind source separation technique based on second order statistics, *IEEE Transactions on Signal Processing* 45 (2) (1997) 434–444.
- [5] J.-F. Cardoso, B.H. Laheld, Equivariant adaptive source separation, *IEEE Transactions on Signal Processing* 44 (12) (1996) 3017–3030.
- [6] A. Cichocki, S.-I. Amari, *Adaptive Blind Signal and Image Processing: Learning Algorithms and Applications*, Wiley, New York, 2002.
- [7] P. Comon, Independent component analysis—a new concept? *Signal Processing* 36 (1994) 287–314.
- [8] A. Hyvärinen, Fast and robust fixed-point algorithms for independent component analysis, *IEEE Transactions on Neural Networks* 10 (3) (1999) 626–634.
- [9] A. Hyvärinen, Complexity pursuit: Separating interesting components from time-series, *Neural Computation* 13 (4) (2001) 883–898.
- [10] A. Hyvärinen, Blind source separation by nonstationarity of variance: A cumulant-based approach, *IEEE Transactions on Neural Networks* 12 (6) (2001) 1471–1474.
- [11] A. Hyvärinen, J. Karhunen, E. Oja, *Independent Component Analysis*, Wiley, New York, 2001.
- [12] C. Jutten, J. Herault, Blind separation of sources, part I: An adaptive algorithm based on neuromimetic architecture, *Signal Processing* 24 (1991) 1–10.
- [13] T.-W. Lee, M. Girolami, T. Sejnowski, Independent component analysis using an extended infomax algorithm for mixed subgaussian and supergaussian sources, *Neural Computation* 11 (2) (1999) 417–441.
- [14] M.S. Lewicki, T.J. Sejnowski, Learning overcomplete representations, *Neural Computation* 12 (2) (2000) 337–365.
- [15] Y.Q. Li, A. Cichocki, S.I. Amari, Analysis of sparse representation and blind source separation, *Neural Computation* 16 (6) (2004) 1193–1234.
- [16] W. Lu, J.C. Rajapakse, Eliminating indeterminacy in ICA, *Neurocomputing* 50 (2003) 271–290.
- [17] D.G. Luenberger, *Optimization by Vector Space Methods*, John Wiley, New York, 1969.
- [18] K. Matsuoka, M. Ohya, M. Kawamoto, A neural net for blind separation of nonstationary signals, *Neural Networks* 8 (3) (1995) 411–419.
- [19] L. Molgedey, H.G. Schuster, Separation of a mixture of independent signals using time delayed correlations, *Physical Review Letters* 72 (23) (1994) 3634–3637.
- [20] E. Oja, M.D. Plumbley, Blind separation of positive sources by globally convergent gradient search, *Neural Computation* 16 (9) (2004) 1811–1825.
- [21] D.-T. Pham, J.-F. Cardoso, Blind separation of instantaneous mixtures of non stationary sources, *IEEE Transactions on Signal Processing* 49 (9) (2001) 1837–1848.
- [22] M.D. Plumbley, E. Oja, A “non-negative PCA” algorithm for independent component analysis, *IEEE Transactions on Neural Networks* 15 (1) (2004) 66–76.
- [23] J.V. Stone, Blind source separation using temporal predictability, *Neural Computation* 13 (2001) 1559–1574.
- [24] Z. Shi, H. Tang, Y. Tang, A new fixed-point algorithm for independent component analysis, *Neurocomputing* 56 (2004) 467–473.
- [25] Z. Shi, H. Tang, Y. Tang, Blind source separation of more sources than mixtures using sparse mixture models, *Pattern Recognition Letters* 26 (16) (2005) 2491–2499.
- [26] Z. Shi, H. Tang, Y. Tang, A fast fixed-point algorithm for complexity pursuit, *Neurocomputing* 64 (2005) 529–536.
- [27] Z. Shi, C. Zhang, Gaussian moments for noisy complexity pursuit, *Neurocomputing* 69 (7–9) (2006) 917–921.
- [28] Z. Shi, C. Zhang, Energy predictability to blind source separation, *Electronics Letters* 42 (17) (2006) 1006–1008.
- [29] Z. Shi, C. Zhang, Semi-blind source extraction for fetal electrocardiogram extraction by combining non-Gaussianity and time-correlation, *Neurocomputing* 70 (2007) 1574–1581.
- [30] Z. Shi, C. Zhang, Nonlinear innovation to blind source separation, *Neurocomputing* 71 (2007) 406–410.
- [31] L. Tong, R.-W. Liu, V. Soon, Y.-F. Huang, Indeterminacy and identifiability of blind identification, *IEEE Transactions on Circuits and Systems* 38 (5) (1991) 499–509.
- [32] R. Vigário, V. Jousmääki, M. Hämäläinen, R. Hari, E. Oja, Independent component analysis for identification of artifacts in magnetoencephalographic recordings, *Advances in Neural Information Processing Systems* 10 (1998) 229–235.
- [33] R. Vigário, J. Särelä, V. Jousmääki, M. Hämäläinen, E. Oja, Independent component approach to the analysis of EEG and MEG recordings, *IEEE Transactions on Biomedical Engineering* 47 (5) (2000) 589–593.
- [34] H. Zhang, C. Guo, Z. Shi, E. Feng, A new constrained fixed-point algorithm for ordering independent components, *Journal of Computational and Applied Mathematics*, in press (doi:10.1016/j.cam.2007.09.010).
- [35] M. Zibulevsky, B.A. Pearlmutter, Blind source separation by sparse decomposition in a signal dictionary, *Neural Computation* 13 (2001) 863–882.