Pattern Recognition 42 (2009) 1732 – 1741

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/pr

Fast nonlinear autocorrelation algorithm for source separation ☆

Zhenwei Shi [a,*], Changshui Zhang [b]

[a] Image Processing Center, School of Astronautics, Beihang University, Beijing 100083, PR China
[b] State Key Laboratory on Intelligent Technology and Systems, Tsinghua National Laboratory for Information Science and Technology (TNList), Department of Automation, Tsinghua University, Beijing 100084, PR China

ARTICLE INFO

Article history:
Received 31 May 2008
Received in revised form 15 November 2008
Accepted 18 December 2008

Keywords:
Blind source separation (BSS)
Independent component analysis (ICA)
Linear autocorrelation
Nonlinear autocorrelation

ABSTRACT

Independent component analysis (ICA) and blind source separation (BSS) methods have been used for pattern recognition problems. It is well known that ICA and BSS depend on the statistical properties of original sources or components, such as non-Gaussianity. In the paper, using a statistical property—nonlinear autocorrelation and maximizing the nonlinear autocorrelation of source signals, we propose a fast fixed-point algorithm for BSS. We study its convergence property and show that its convergence speed is at least quadratic. Simulations by the artificial signals and the real-world applications verify the efficient implementation of the proposed method.

© 2008 Elsevier Ltd. All rights reserved.

1. Introduction

Independent component analysis (ICA) or blind source separation (BSS) [6,11] is an increasingly popular data analysis technique which has received wide attention in various fields such as pattern recognition, biomedical signal processing and analysis, speech and image processing, wireless telecommunication systems and data mining. The main task of ICA or BSS is to recover original sources from their mixtures using some statistical properties of original sources. In fact, ICA can be seen as a method for solving the BSS, which assumes that the sources are non-Gaussian.

Several methods for BSS using the statistical properties of original sources have been proposed by researchers in applied mathematics, neural networks, pattern recognition and statistical signal processing, such as non-Gaussianity (or equivalently, ICA) [1,3,5–8,11–13,16], linear predictability or smoothness [2,6], linear autocorrelation [4,18,31], coding complexity [10,24,25,27], temporal predictability [30], energy predictability [26], nonstationarity [9,17,21], sparsity [14,15,23,34], nonnegativity [20,22]and nonlinear innovation [28].

In our previous work, we have presented a new way using the nonlinear autocorrelation of source signals for BSS [29]. However, the convergence speed of the algorithm used there is slow, in fact, its convergence speed can be shown to be linear. In this paper, a fast and robust fixed-point algorithm using the nonlinear autocorrelation is proposed. Further, we study its convergence property and the convergence speed is shown to be at least quadratic. Also, the fast nonlinear autocorrelation BSS method is demonstrated in detail by simulations. Simulations on the artificial signals with square temporal autocorrelation, artificial electrocardiograph (ECG) signals, the real-world ECG data and the real-world magnetoencephalography (MEG) data verify the efficient implementation of the proposed method.

2. Nonlinear autocorrelation to BSS

2.1. Objective function

We observe sensor signals or data $\mathbf{x}(t)=(x_1(t),\ldots,x_n(t))^{\mathrm{T}}$ described by the matrix equation:

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

where $\mathbf{A}$ is an $n\times n$ unknown mixing matrix, $\mathbf{s}(t)=(s_1(t),\ldots,s_n(t))^{\mathrm{T}}$ are unknown original sources, which are assumed to be zero-mean and unit-variance. We assume that original sources are mutually independent and have the nonlinear autocorrelation (see below).

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

$$y(t)=\mathbf{w}^{\mathrm{T}}\mathbf{x}(t), \tag{2}$$

$$y(t-\tau)=\mathbf{w}^{\mathrm{T}}\mathbf{x}(t-\tau), \tag{3}$$

where $\mathbf{w}=(w_1,\ldots,w_n)^{\mathrm{T}}$ is the weight vector, and τ is some lag constant, often equal to 1.

☆ The work was supported by the National Natural Science Foundation of China under Grants 60605002 and 60835002, and the Research Fund for the Doctoral Program of Higher Education of China under Grant 200800061056.

* Corresponding author. Tel.: +86 10 823 16 502; fax: +86 10 627 86 911.
E-mail addresses: shizhenwei@mail.tsinghua.edu.cn (Z. Shi), zcs@mail.tsinghua.edu.cn (C. Zhang).

0031-3203/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.patcog.2008.12.025

We assume that the measured sensor signals $\mathbf{x}$ have already been followed by an $n \times n$ whitening matrix $\mathbf{V}$ such that the components of $\tilde{\mathbf{x}}(t) = \mathbf{V}\mathbf{x}(t)$ are unit variance and uncorrelated. We present the following constrained maximization problem based on the nonlinear autocorrelation of the desired source [29]:

$$\begin{aligned}\max_{\|\mathbf{w}\|=1} \Psi(\mathbf{w}) &= E\{G(\tilde{y}(t))G(\tilde{y}(t-\tau))\} \\ &= E\{G(\mathbf{w}^{\mathrm{T}}\tilde{\mathbf{x}}(t))G(\mathbf{w}^{\mathrm{T}}\tilde{\mathbf{x}}(t-\tau))\},\end{aligned} \tag{4}$$

where G is a differentiable nonlinear function, which measures the nonlinear autocorrelation degree of the desired source. Examples of choices of G are $G(u) = u^2$ and $G(u) = \log\cosh(u)$. In contrast with the well-known linear autocorrelation (where G is a linear function) for BSS, such as the methods [4,18,31], the nonlinear autocorrelation is a novel statistical property for BSS.

2.2. Learning algorithm

Maximizing the objective function in (4), we can derive a fast fixed-point algorithm through the approximative Newton iteration, which is similar to the well-known FastICA [8] in ICA. The gradient of $E\{G(\tilde{y}(t))G(\tilde{y}(t-\tau))\}$ with respect to $\mathbf{w}$ can be obtained as

$$\begin{aligned}\frac{\partial E\{G(\tilde{y}(t))G(\tilde{y}(t-\tau))\}}{\partial \mathbf{w}} &= E\{g(\tilde{y}(t))G(\tilde{y}(t-\tau))\tilde{\mathbf{x}}(t) \\ &\quad + G(\tilde{y}(t))g(\tilde{y}(t-\tau))\tilde{\mathbf{x}}(t-\tau)\},\end{aligned} \tag{5}$$

where the function g is the derivative of G. The Hessian matrix of $E\{G(\tilde{y}(t))G(\tilde{y}(t-\tau))\}$ with respect to $\mathbf{w}$ can be obtained as

$$\begin{aligned}\frac{\partial^2 E\{G(\tilde{y}(t))G(\tilde{y}(t-\tau))\}}{\partial \mathbf{w}^2} &= E\{g'(\tilde{y}(t))G(\tilde{y}(t-\tau))\tilde{\mathbf{x}}(t)\tilde{\mathbf{x}}(t)^{\mathrm{T}} \\ &\quad + G(\tilde{y}(t))g'(\tilde{y}(t-\tau))\tilde{\mathbf{x}}(t-\tau)\tilde{\mathbf{x}}(t-\tau)^{\mathrm{T}} \\ &\quad + g(\tilde{y}(t))g(\tilde{y}(t-\tau))\tilde{\mathbf{x}}(t)\tilde{\mathbf{x}}(t-\tau)^{\mathrm{T}} \\ &\quad + g(\tilde{y}(t))g(\tilde{y}(t-\tau))\tilde{\mathbf{x}}(t-\tau)\tilde{\mathbf{x}}(t)^{\mathrm{T}}\},\end{aligned} \tag{6}$$

where the function g' is the derivative of g.

Thus, we have the following Newton iteration:

$$\mathbf{w} \leftarrow \mathbf{w} - \left(\frac{\partial^2 E\{G(\tilde{y}(t))G(\tilde{y}(t-\tau))\}}{\partial \mathbf{w}^2}\right)^{-1} \left(\frac{\partial E\{G(\tilde{y}(t))G(\tilde{y}(t-\tau))\}}{\partial \mathbf{w}}\right), \tag{7}$$

$$\mathbf{w} \leftarrow \mathbf{w}/\|\mathbf{w}\|. \tag{8}$$

The Newton algorithm requires the inverse of the Hessian matrix, which is generally difficult to compute. However, based on the BSS assumptions, we can simplify the Hessian matrix to obtain a simple fixed-point algorithm.

Assume that $\mathbf{x}$ and $\mathbf{s}$ follow the BSS mixing model: $\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t)$. After data $\mathbf{x}(t)$ are whitened, we have $\tilde{\mathbf{x}}(t) = \tilde{\mathbf{A}}\mathbf{s}(t)$, where the new mixing matrix $\tilde{\mathbf{A}} = \mathbf{V}\mathbf{A}$ is orthogonal [11]. We obtain

$$E\{\tilde{\mathbf{x}}(t)\tilde{\mathbf{x}}(t)^{\mathrm{T}}\} = \tilde{\mathbf{A}}E\{\mathbf{s}(t)\mathbf{s}(t)^{\mathrm{T}}\}\tilde{\mathbf{A}}^{\mathrm{T}} = \tilde{\mathbf{A}}\tilde{\mathbf{A}}^{\mathrm{T}} = \mathbf{I}, \tag{9}$$

$$E\{\tilde{\mathbf{x}}(t)\tilde{\mathbf{x}}(t-\tau)^{\mathrm{T}}\} = E\{\tilde{\mathbf{x}}(t-\tau)\tilde{\mathbf{x}}(t)^{\mathrm{T}}\} = \tilde{\mathbf{A}}E\{\mathbf{s}(t)\mathbf{s}(t-\tau)^{\mathrm{T}}\}\tilde{\mathbf{A}}^{\mathrm{T}} = \tilde{\mathbf{A}}\tilde{\mathbf{A}}^{\mathrm{T}} = \mathbf{I}, \tag{10}$$

where we assume that $E\{\mathbf{s}(t)\mathbf{s}(t)^{\mathrm{T}}\} = E\{\mathbf{s}(t)\mathbf{s}(t-\tau)^{\mathrm{T}}\} = \mathbf{I}$. Since the data are whitened, the reasonable approximations seem to be

$$\begin{aligned}&E\{g'(\tilde{y}(t))G(\tilde{y}(t-\tau))\tilde{\mathbf{x}}(t)\tilde{\mathbf{x}}(t)^{\mathrm{T}} + G(\tilde{y}(t))g'(\tilde{y}(t-\tau))\tilde{\mathbf{x}}(t-\tau)\tilde{\mathbf{x}}(t-\tau)^{\mathrm{T}}\} \\ &\quad \approx E\{g'(\tilde{y}(t))G(\tilde{y}(t-\tau))\}E\{\tilde{\mathbf{x}}(t)\tilde{\mathbf{x}}(t)^{\mathrm{T}}\} \\ &\qquad + E\{G(\tilde{y}(t))g'(\tilde{y}(t-\tau))\}E\{\tilde{\mathbf{x}}(t-\tau)\tilde{\mathbf{x}}(t-\tau)^{\mathrm{T}}\} \\ &\quad = E\{g'(\tilde{y}(t))G(\tilde{y}(t-\tau)) + G(\tilde{y}(t))g'(\tilde{y}(t-\tau))\}\mathbf{I}\end{aligned} \tag{11}$$

and

$$\begin{aligned}&E\{g(\tilde{y}(t))g(\tilde{y}(t-\tau))\tilde{\mathbf{x}}(t)\tilde{\mathbf{x}}(t-\tau)^{\mathrm{T}} + g(\tilde{y}(t))g(\tilde{y}(t-\tau))\tilde{\mathbf{x}}(t-\tau)\tilde{\mathbf{x}}(t)^{\mathrm{T}}\} \\ &\quad \approx E\{g(\tilde{y}(t))g(\tilde{y}(t-\tau))\}E\{\tilde{\mathbf{x}}(t)\tilde{\mathbf{x}}(t-\tau)^{\mathrm{T}}\} \\ &\qquad + E\{g(\tilde{y}(t))g(\tilde{y}(t-\tau))\}E\{\tilde{\mathbf{x}}(t-\tau)\tilde{\mathbf{x}}(t)^{\mathrm{T}}\} \\ &\quad = 2E\{g(\tilde{y}(t))g(\tilde{y}(t-\tau))\}\mathbf{I}.\end{aligned} \tag{12}$$

Thus, the Hessian matrix (6) becomes diagonal, i.e., $\partial^2 E\{G(\tilde{y}(t))G(\tilde{y}(t-\tau))\}/\partial \mathbf{w}^2 \approx \beta\mathbf{I}$ (where $\beta = E\{g'(\tilde{y}(t))G(\tilde{y}(t-\tau)) + G(\tilde{y}(t))g'(\tilde{y}(t-\tau))\} + 2E\{g(\tilde{y}(t))g(\tilde{y}(t-\tau))\}$) and can easily be inverted.

As a result, formula (7) can be further simplified by multiplying both sides of (7) by $\beta\mathbf{I}$. This gives, after the straightforward algebraic simplification, an approximative Newton iteration:

$$\mathbf{w} \leftarrow \frac{\partial E\{G(\tilde{y}(t))G(\tilde{y}(t-\tau))\}}{\partial \mathbf{w}} - \frac{\partial^2 E\{G(\tilde{y}(t))G(\tilde{y}(t-\tau))\}}{\partial \mathbf{w}^2}\mathbf{w}, \tag{13}$$

$$\mathbf{w} \leftarrow \mathbf{w}/\|\mathbf{w}\|. \tag{14}$$

Thus, we obtain the following fixed-point algorithm:

$$\begin{aligned}\mathbf{w} \leftarrow{}& E\{g(\tilde{y}(t))G(\tilde{y}(t-\tau))\tilde{\mathbf{x}}(t) + G(\tilde{y}(t))g(\tilde{y}(t-\tau))\tilde{\mathbf{x}}(t-\tau)\} \\ &- E\{g'(\tilde{y}(t))G(\tilde{y}(t-\tau)) + G(\tilde{y}(t))g'(\tilde{y}(t-\tau))\}\mathbf{w} \\ &- E\{g(\tilde{y}(t))g(\tilde{y}(t-\tau))\}(E\{\tilde{\mathbf{x}}(t)\tilde{\mathbf{x}}(t-\tau)^{\mathrm{T}}\} \\ &+ E\{\tilde{\mathbf{x}}(t-\tau)\tilde{\mathbf{x}}(t)^{\mathrm{T}}\})\mathbf{w}, \mathbf{w} \leftarrow \mathbf{w}/\|\mathbf{w}\|.\end{aligned} \tag{15}$$

Note that convergence of the fixed-point algorithm means that the old and new values of $\mathbf{w}$ point in the same direction [11].

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

2.3. Convergence analysis of the proposed algorithm

In this section, we obtain the convergence properties of the fixed-point algorithm (15).[1] We have the following theorem:

Theorem 1. *Assume that the following conditions are satisfied* ($\forall j \neq i$): (a) $\{s_i, s_{i\tau}\}$ *and* $\{s_j, s_{j\tau}\}$ *are mutually independent*; (b) $E\{s_l s_{l\tau}\} = 0$ ($\forall l$); (c) $E\{s_i g(s_i)G(s_{i\tau}) + s_{i\tau}g(s_{i\tau})G(s_i) - g'(s_i)G(s_{i\tau}) - g'(s_{i\tau})G(s_i)\} \neq 0$. *Then, the fixed-point algorithm* (15) *converges. This means that vector* $\mathbf{w}$ *converges, up to the sign, to one row of the inverse of the mixing matrix* $\mathbf{VA}$, *and the convergence speed is at least quadratic.*

Proof. To begin with, we make the change of variable $\mathbf{p} = (p_1, \ldots, p_n)^{\mathrm{T}} = \mathbf{A}^{\mathrm{T}}\mathbf{V}^{\mathrm{T}}\mathbf{w}$. From $E\{\mathbf{s}\mathbf{s}^{\mathrm{T}}\} = \mathbf{I}$ and the property of whitening, the fixed-point algorithm (15) is given by

$$\begin{aligned}\hat{\mathbf{p}} = {}& E\{\mathbf{s}g(\mathbf{p}^{\mathrm{T}}\mathbf{s})G(\mathbf{p}^{\mathrm{T}}\mathbf{s}_\tau) + \mathbf{s}_\tau g(\mathbf{p}^{\mathrm{T}}\mathbf{s}_\tau)G(\mathbf{p}^{\mathrm{T}}\mathbf{s})\} \\ &- E\{g'(\mathbf{p}^{\mathrm{T}}\mathbf{s})G(\mathbf{p}^{\mathrm{T}}\mathbf{s}_\tau) + g'(\mathbf{p}^{\mathrm{T}}\mathbf{s}_\tau)G(\mathbf{p}^{\mathrm{T}}\mathbf{s})\}\mathbf{p} \\ &- E\{(\mathbf{s}\mathbf{s}_\tau^{\mathrm{T}} + \mathbf{s}_\tau\mathbf{s}^{\mathrm{T}})g(\mathbf{p}^{\mathrm{T}}\mathbf{s})g(\mathbf{p}^{\mathrm{T}}\mathbf{s}_\tau)\}\mathbf{p},\end{aligned} \tag{16}$$

$$\mathbf{p} = \hat{\mathbf{p}}/\|\hat{\mathbf{p}}\|. \tag{17}$$

Using a Taylor approximation for G, g and g', we have

$$\begin{aligned}G(\mathbf{p}^{\mathrm{T}}\mathbf{s}) = {}& G(p_1 s_1) + g(p_1 s_1)\mathbf{p}_{-1}^{\mathrm{T}}\mathbf{s}_{-1} \\ &+ \tfrac{1}{2}g'(p_1 s_1)(\mathbf{p}_{-1}^{\mathrm{T}}\mathbf{s}_{-1})^2 + O(\|\mathbf{p}_{-1}\|^3),\end{aligned} \tag{18}$$

$$\begin{aligned}g(\mathbf{p}^{\mathrm{T}}\mathbf{s}) = {}& g(p_1 s_1) + g'(p_1 s_1)\mathbf{p}_{-1}^{\mathrm{T}}\mathbf{s}_{-1} \\ &+ \tfrac{1}{2}g''(p_1 s_1)(\mathbf{p}_{-1}^{\mathrm{T}}\mathbf{s}_{-1})^2 + O(\|\mathbf{p}_{-1}\|^3),\end{aligned} \tag{19}$$

$$\begin{aligned}g'(\mathbf{p}^{\mathrm{T}}\mathbf{s}) = {}& g'(p_1 s_1) + g''(p_1 s_1)\mathbf{p}_{-1}^{\mathrm{T}}\mathbf{s}_{-1} \\ &+ \tfrac{1}{2}g'''(p_1 s_1)(\mathbf{p}_{-1}^{\mathrm{T}}\mathbf{s}_{-1})^2 + O(\|\mathbf{p}_{-1}\|^3),\end{aligned} \tag{20}$$

[1] We will drop the sampling index t for simplicity, i.e., $\mathbf{s}_\tau = \mathbf{s}(t-\tau)$ and $s_{i\tau} = s_i(t-\tau)$. For source signals, we assume that $E\{s_l^2\} = E\{s_{l\tau}^2\} = 1 (\forall l)$.

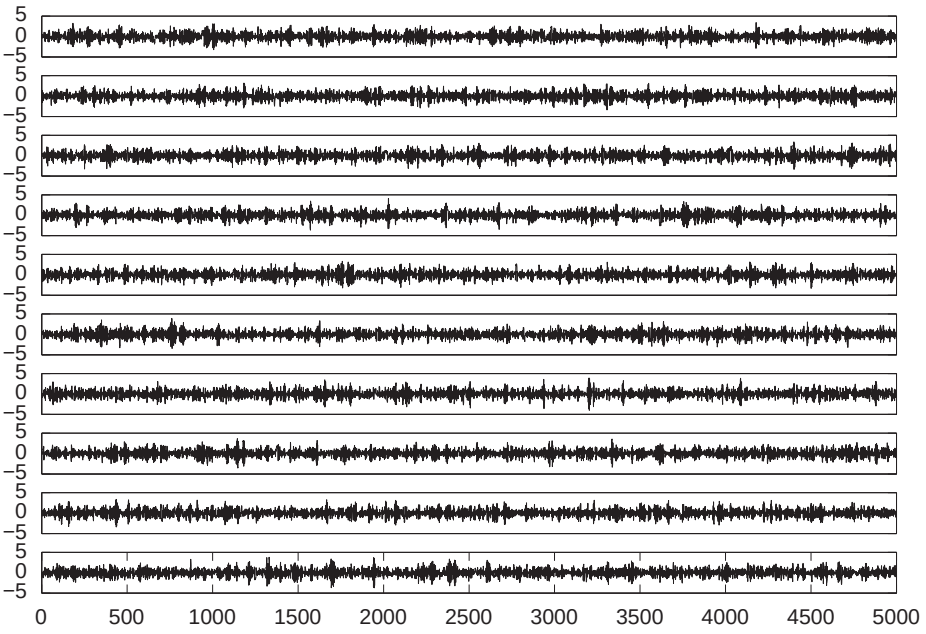

Fig. 1. Ten sources with square temporal autocorrelation.

where $\mathbf{p}_{-1}$ and $\mathbf{s}_{-1}$ are the vectors $\mathbf{p}$ and $\mathbf{s}$ without their first components. We analyze the convergence of the point $\mathbf{p}=\mathbf{e}$, where $\mathbf{e}_1=(1,0,0,\dots)^{\mathrm{T}}$. Using assumptions (a) and (b) in Theorem 1, then doing some algebraic manipulations, we obtain

$$\hat{p}_1 = E\{s_1 g(s_1)G(s_{1\tau}) + s_{1\tau}g(s_{1\tau})G(s_1) - g'(s_1)G(s_{1\tau}) - g'(s_{1\tau})G(s_1)\} + O(\|\mathbf{p}_{-1}\|^3), \tag{21}$$

$$\hat{p}_i = E\{s_i^2 s_{i\tau}g'(s_1)g(s_{1\tau}) + \tfrac{1}{2}s_i^3 g''(s_1)G(s_{1\tau}) + \tfrac{1}{2}s_i s_{i\tau}^2 g(s_1)g'(s_{1\tau})\}p_i^2 + E\{s_{i\tau}^2 s_i g'(s_{1\tau})g(s_1) + \tfrac{1}{2}s_{i\tau}^3 g''(s_{1\tau})G(s_1) + \tfrac{1}{2}s_{i\tau}s_i^2 g(s_{1\tau})g'(s_1)\}p_i^2 + O(\|\mathbf{p}_{-1}\|^3). \tag{22}$$

This shows clearly that under the assumption $E\{s_1 g(s_1)G(s_{1\tau}) + s_{1\tau}g(s_{1\tau})G(s_1) - g'(s_1)G(s_{1\tau}) - g'(s_{1\tau})G(s_1)\}\neq 0$, algorithm (15) converges to such a vector $\mathbf{e}$ that $e_1=\pm 1$ and $e_j=0$, $\forall j\neq 1$. This implies that $\mathbf{w}$ converges as Theorem 1 stated. Eq. (22) shows that the convergence is at least quadratic. □

From the convergence theorem, we know that the fixed-point algorithm (15) optimizes the objective function (4) very fast, thus, we refer to this algorithm as fast algorithm for nonlinear autocorrelation (FastNA).

3. Experimental results

3.1. Experiments on artificial signals

We demonstrate when we choose $G(u)=u^2$ and $G(u)=\log(\cosh(u))$, the FastNA algorithms can separate the source signals with square temporal autocorrelation.

We create 10 artificial signals which have square temporal autocorrelation as follows (with Gaussian marginal distribution, zero linear autocorrelation)[2] [9]: first, we create 10 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 had identical autoregressive coefficients (0.9). 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 10 source signals are shown in Fig. 1. The source signals are mixed with 10×10 random matrices. The FastNA algorithms ($\tau=1$, $G(u)=u^2, G(u)=\log(\cosh(u))$) 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) ($\tau=1$) [9] and the fixed-point algorithm for BSS using the nonlinear autocorrelation (FixNA) with $G(u)=u^2$ and $G(u)=\log(\cosh(u))$ ($\tau=1$) [29]. In order to measure the accuracy of separation, we calculate the performance index [1,26,28]

$$\mathbf{PI}=\frac{1}{n^2}\left\{\sum_{i=1}^{n} r\mathbf{PI}_i+\sum_{j=1}^{n} c\mathbf{PI}_j\right\} =\frac{1}{n^2}\left\{\sum_{i=1}^{n}\left(\sum_{j=1}^{n}\frac{|p_{ij}|}{\max_k|p_{ik}|}-1\right)+\sum_{j=1}^{n}\left(\sum_{i=1}^{n}\frac{|p_{ij}|}{\max_k|p_{kj}|}-1\right)\right\}, \tag{23}$$

where $r\mathbf{PI}_i=\sum_{j=1}^{n}|p_{ij}|/\max_k|p_{ik}|-1$ and $c\mathbf{PI}_j=\sum_{i=1}^{n}|p_{ij}|/\max_k|p_{kj}|-1$ in which p_{ij} is the ijth 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 **PI** is, the poorer the statistical performance of a BSS algorithm [1,26,28].

Fig. 2 shows the average performance indexes over 100 independent trials against iteration numbers by the five algorithms. The FastNA algorithm using $G(u)=u^2$ performs similarly to the FPNSV and the FixNA using $G(u)=u^2$ in the sense of the separation accuracy. The FastNA algorithm using $G(u)=\log(\cosh(u))$ and the FixNA algorithm using $G(u)=\log(\cosh(u))$ have a slightly better performance than the other three algorithms. Also it is worth noting that the two FastNA algorithms and the FPNSV algorithm converge very fast, only 6–7 iterations were necessary, on the average, to achieve the maximum accuracy. The fact illustrates the fast convergence of the proposed fixed-point algorithms as shown in Theorem 1. However, the two FixNA algorithms converge slower than the other three algorithms.

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

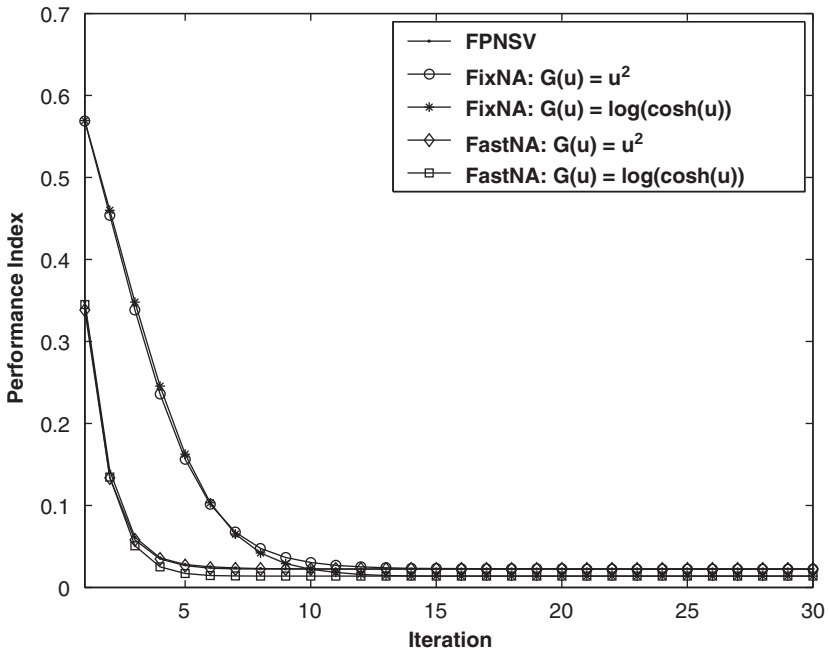

Fig. 2. Average performance indexes over 100 independent runs for 10 sources with square temporal autocorrelation by the different algorithms.

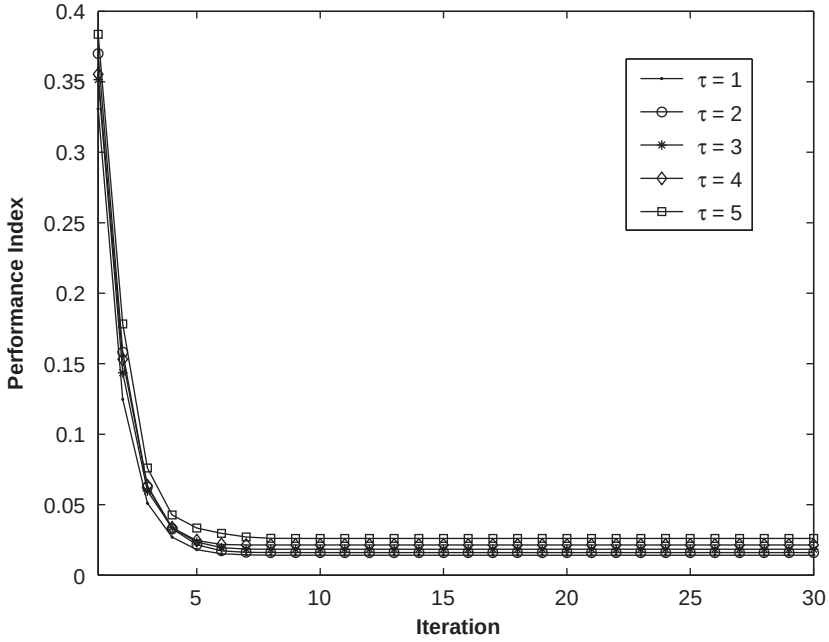

Fig. 3. The average performance indexes over 100 independent trials against iteration numbers by the proposed FastNA algorithm with the different time lag τ.

To investigate the robustness of FastNA algorithm for choosing the different time lag τ, Fig. 3 shows the average performance indexes over 100 independent trials against iteration numbers by the proposed FastNA algorithm ($G(u) = \log(\cosh(u))$) with the different time lag τ ($\tau = 1$, 2, 3, 4, and 5, respectively). We can see that the performance of the FastNA algorithm with $\tau = 1$ is best, however, the performance with other lag τ (whose value is close to $\tau = 1$) is also similar to $\tau = 1$. Thus, from the experiments, we can see that the performance of the proposed FastNA is still good if the choice of the time delay τ is not far from the time delay $\tau = 1$. However, if the time delay τ is far from the true one, the performance of the FastNA degrades.

To investigate the robustness of the algorithms further, we randomly add different outliers whose values are 10 in each source signal. Figs. 4–6 show the average performance indexes over 100 independent trials against iteration numbers when the outliers are 20, 25 and 30 ($\tau = 1$). Obviously, the FastNA algorithm using $G(u) = \log(\cosh(u))$ and the FixNA algorithm using $G(u) = \log(\cosh(u))$ outperform the other three algorithms when the outliers are introduced. In fact, the FPNSV algorithm, the FastNA using $G(u) = u^2$, and the FixNA using $G(u) = u^2$ are all sensitive to the outliers and degrade in performance. Further, the FastNA algorithm using $G(u) = \log(\cosh(u))$ converges faster than the FixNA algorithm using $G(u) = \log(\cosh(u))$. Thus, we can see that the nonlinear $G(u) = \log(\cosh(u))$ could have

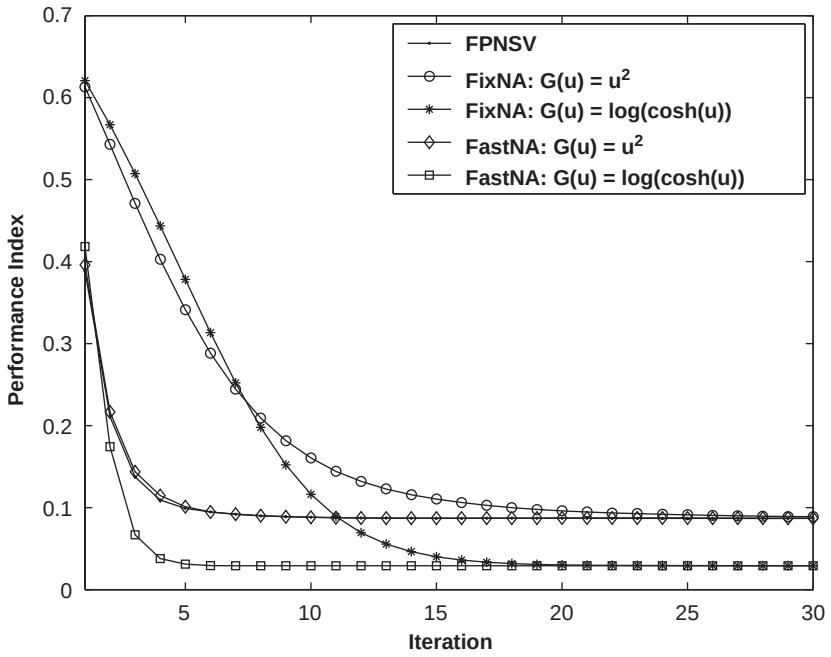

Fig. 4. Average performance indexes over 100 independent runs for 10 sources with square temporal autocorrelation when 20 outliers are introduced.

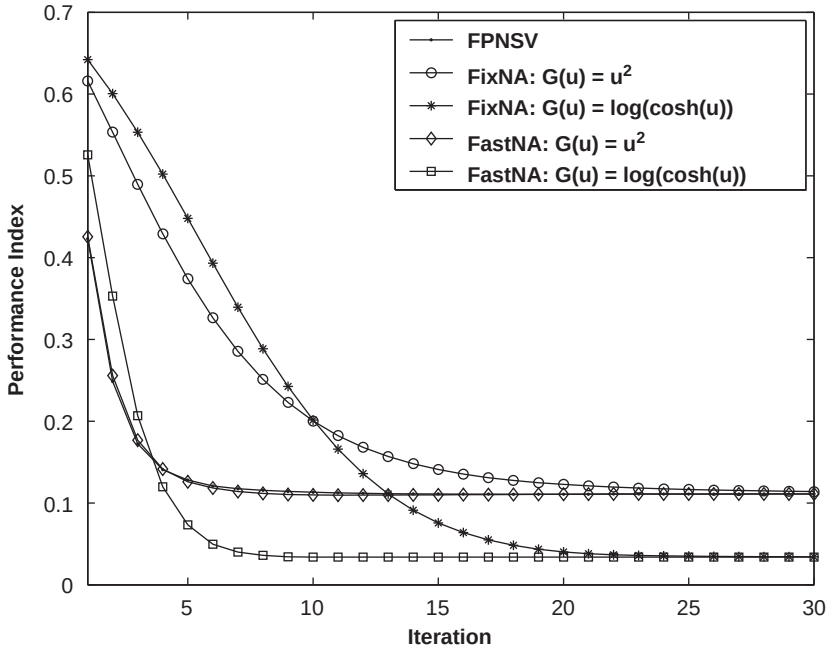

Fig. 5. Average performance indexes over 100 independent runs for 10 sources with square temporal autocorrelation when 25 outliers are introduced.

better convergence in practice and the FastNA algorithm using $G(u)=\log(\cosh(u))$ performs best.

In fact, in the FPNSV algorithm, the fourth-order cross-cumulant could be considered as a normalized version of the autocorrelation of square (energy) [9]. In particular, if the signals have no linear time-correlation, and $E\{s_l^2\}=E\{s_{l,\tau}^2\}=1$ $(\forall l)$, the cross-cumulant is equal to the autocorrelation of the square (energy) as the objective function of FastNA when the nonlinear $G(u)=u^2$ is chosen. This is the reason why the two algorithms perform similarly in the sense of the separation accuracy.

On the other hand, the FPNSV is also a fixed-point algorithm and the convergence of the algorithm can be proven to be cubic [9]. The optimization problem which leads to the FPNSV algorithm is formally identical to the one encountered when the absolute value of kurtosis is maximized to find the most non-Gaussian directions in ICA. The only difference is that the kurtoses of the source signals are replaced by the cumulants measuring nonstationarity [9]. Thus, we can believe that the FPNSV is similar to the well-known kurtosis-based fixed-point FastICA which is extended to nonstationary sources. Furthermore, the performance of the FPNSV for nonstationary sources is similar to that of the kurtosis-based FastICA for non-Gaussian sources, i.e., the FPNSV is also sensitive to the outliers as the kurtosis-based FastICA (so are the FastNA and FixNA algorithms using $G(u)=u^2$). However, the nonlinear function $G(u)=\log(\cosh(u))$

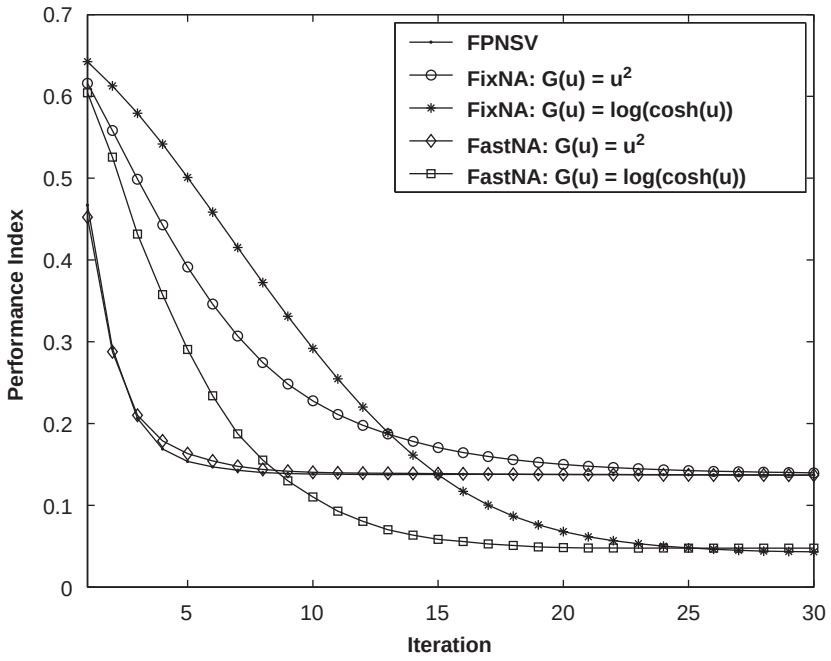

Fig. 6. Average performance indexes over 100 independent runs for 10 sources with square temporal autocorrelation when 30 outliers are introduced.

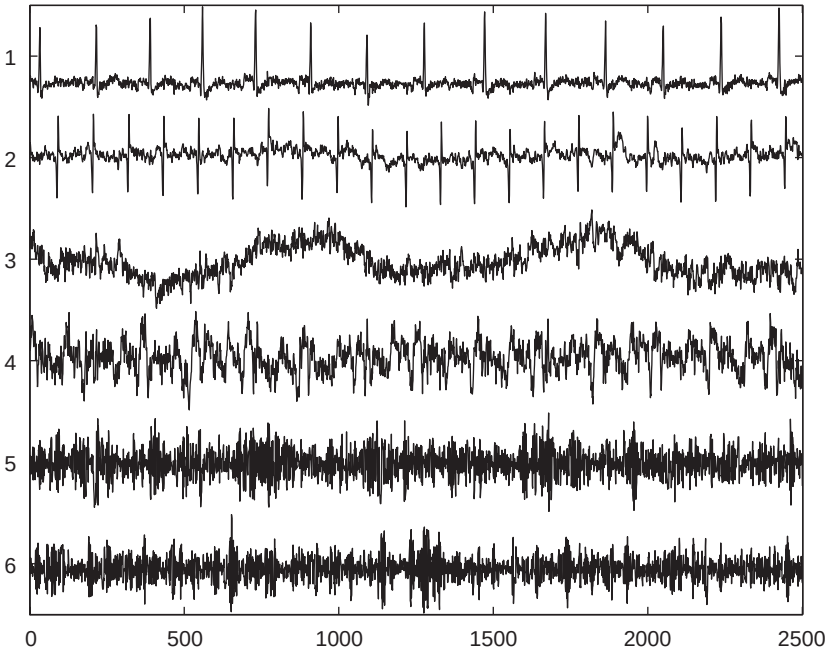

Fig. 7. Six artificial ECG signals used in the simulations. (1) MECG. (2) FECG. (3) Breathing artifact. (4) Electrode artifact. (5 and 6) Two square temporal autocorrelation Gaussian signals.

used by the FastNA is better than the kurtosis-based function, due to its better analysis property and robustness to outliers [11].

3.2. Experiments on ECG data

The separation or extraction of fetal electrocardiogram (FECG) using noninvasive techniques is an important challenge in biomedical signal processing and pattern recognition. The FECG contains important information about the health and condition of the fetus. However, FECG is always corrupted by various kinds of noise, such as the maternal electrocardiogram (MECG) with extreme high amplitude, respiration and stomach activity, thermal noise and noise from electrode-skin contact. ICA and blind source extraction (BSE) methods have been used for the ECG data to extract the artifacts [2,27].

Unlike ICA which assumes that the source signals are non-Gaussian, we use the nonlinear autocorrelation to capture the statistical character of components. In the following experiments, we use the artificial ECG data and the real-world ECG data to verify the efficient implementation of the proposed method.

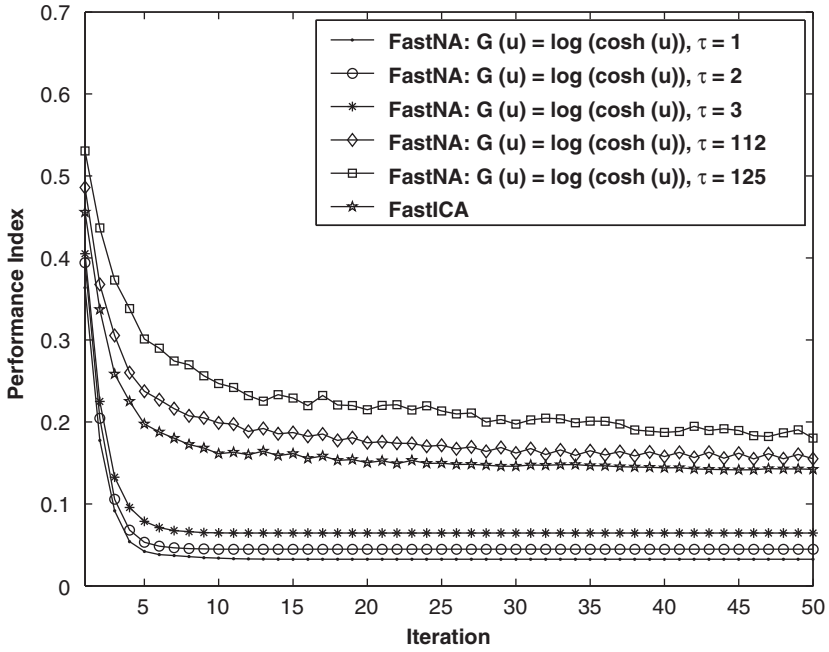

Fig. 8. Average performance indexes over 100 independent runs.

3.2.1. Experiments on artificial ECG data

In the following, we make some experiments on artificial ECG data in order to separate the clearer FECG and MECG signals. We adopt six zero-mean and unit-variance source signals (2500 samples), shown in Fig. 7. From the top to down, they are, an MECG, an FECG (whose period is 112), a breathing artifact, an electrode artifact and two square temporal autocorrelation Gaussian signals adopted from the above experiment. The observed signals are generated by a 6×6 random mixing matrix. In the experiments, we compare the well-known FastICA algorithm [8] and the proposed FastNA algorithms ($G(u)=\log(\cosh(u))$) with the different time lag τ ($\tau=1$, 2, 3, 112 and 125, respectively). Fig. 8 shows comparison results of the average performance indexes values of 100 independent trials by the six algorithms. We can see that the performance of the FastNA with $\tau=1$ is best. The performances of the FastNA with the lags $\tau=2$ and 3 are close to the FastNA using $\tau=1$, however, are not as good as the FastNA using $\tau=1$. On the other hand, the FastNA with $\tau=112$ and 125 and the FastICA perform degrade. In fact, the three algorithms fail to these data. The reason may be that the nonlinear autocorrelation values of some signals are small at some lags (such as $\tau=112$ and 125) and the FastICA cannot separate two Gaussian signals. Thus, we can adopt the FastNA algorithm with $\tau=1$ in practice.

3.2.2. Experiments on real-world ECG data

We perform further experiments on real-world ECG data—the well-known ECG measured from a pregnant woman and distributed by De Moor [19]. Only 10 s of recordings (resampled at 250 Hz)[3] are displayed in Fig. 9.

The output components of FastNA using the nonlinear $G(u)=\log(\cosh(u))$ with the deflation scheme ($\tau=1$) are shown in Fig. 10. The success of the decomposition is already seen by the visual inspection. Obviously, channels 1 and 2 are dominated by the heartbeat of the mother, and channel 3 by that of the child. Other channels still contain heartbeat components (of mother and child), but look much more noisy. In fact, the extraction results of the MECG and FECG signals are similar to the well-known ICA (some other output components of the FastNA are different from ICA). Thus, the nonlinear temporal autocorrelation provides an alternative statistical property for the FECG and it is also likely a strong feature of ECG data sets. Furthermore, the proposed FastNA algorithm provides a fast and robust BSS method when the square temporal autocorrelation Gaussian signals appear.

3.3. Applications in real-world MEG data analysis

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 using an MEG record, as a research or clinical tool, the investigator may face a problem of extracting the essential features of the neuromagnetic signals in the presence of artifacts. The amplitude of the disturbance may be higher than that of the brain signals, and the artifacts may resemble pathological signals in shape. The identification and eventual removal of artifacts is a challenging problem in MEG [32,33].

ICA has been shown to extract and eliminate the artifacts from the MEG signals [32,33]. However, unlike ICA which assumes that the source signals are non-Gaussian, we use the nonlinear autocorrelation to capture the statistical character of sources. 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 17 730 samples.

Fig. 11 shows the extracted 20 nonlinear autocorrelation source signals using the proposed FastNA approach (using $G(u)=\log(\cosh(u))$ with the deflation scheme). It is known that some of extracted source signals have clear physical or physiological interpretation [32,33], i.e., the extracted source signals 1 and 3 correspond to eye movements, signal 6 corresponds to the heart

[3] Although in De Moor's homepage he assured that the sampling frequency was 500 Hz, Barros et al. believed that it was most likely to be 250 Hz [2].

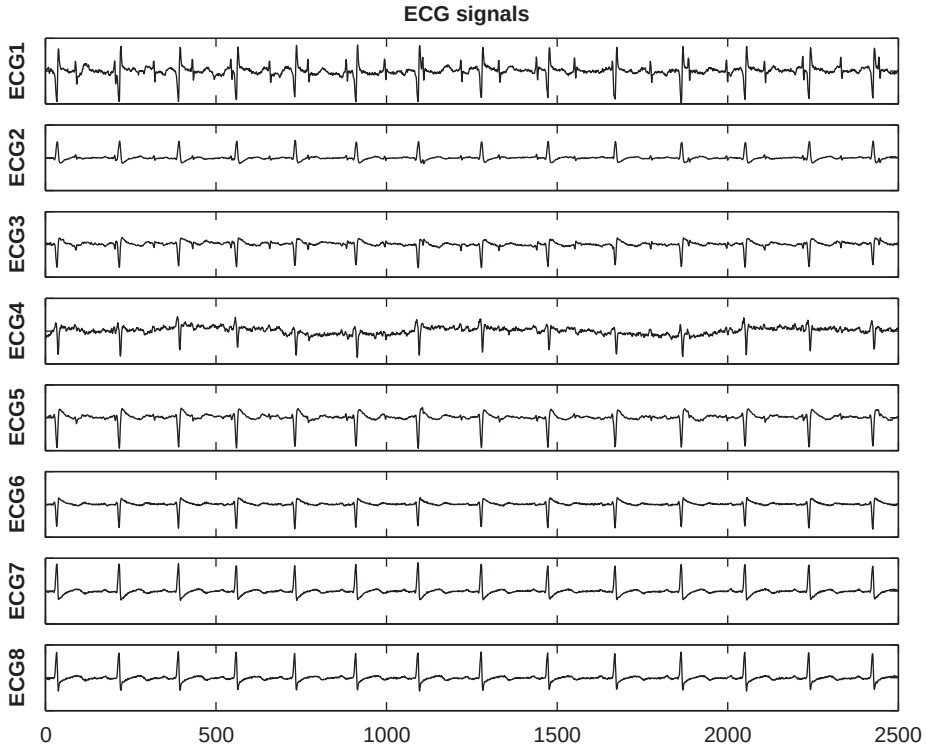

Fig. 9. ECG signals measured from a pregnant woman.

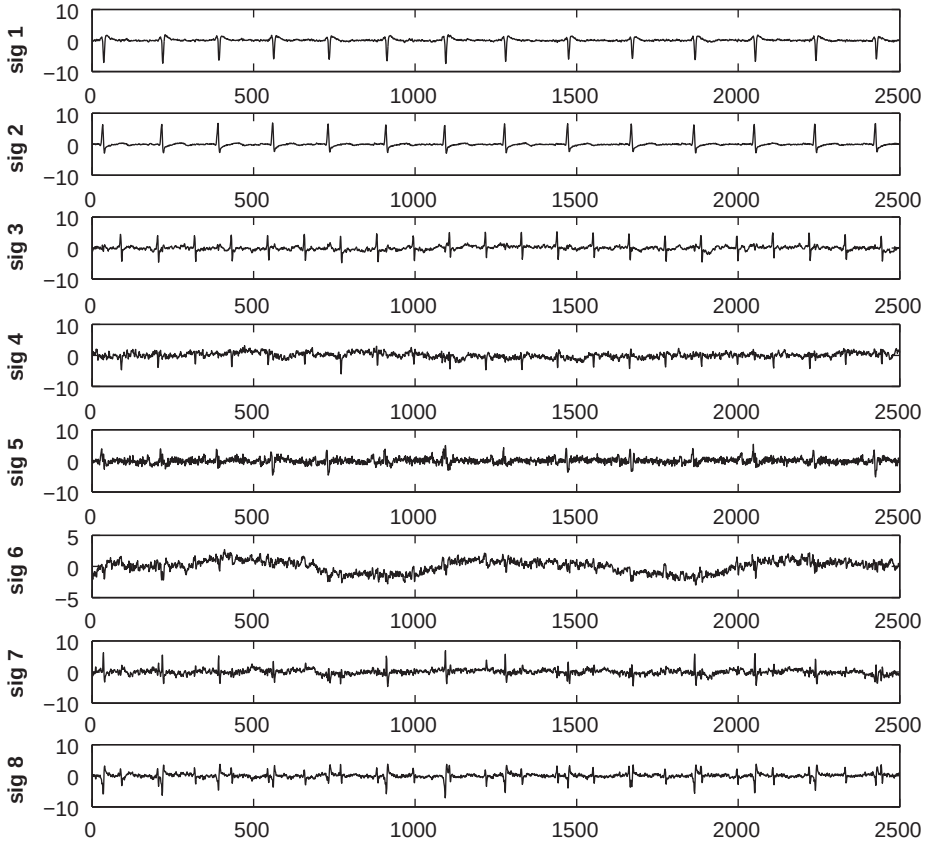

Fig. 10. The output components of FastNA.

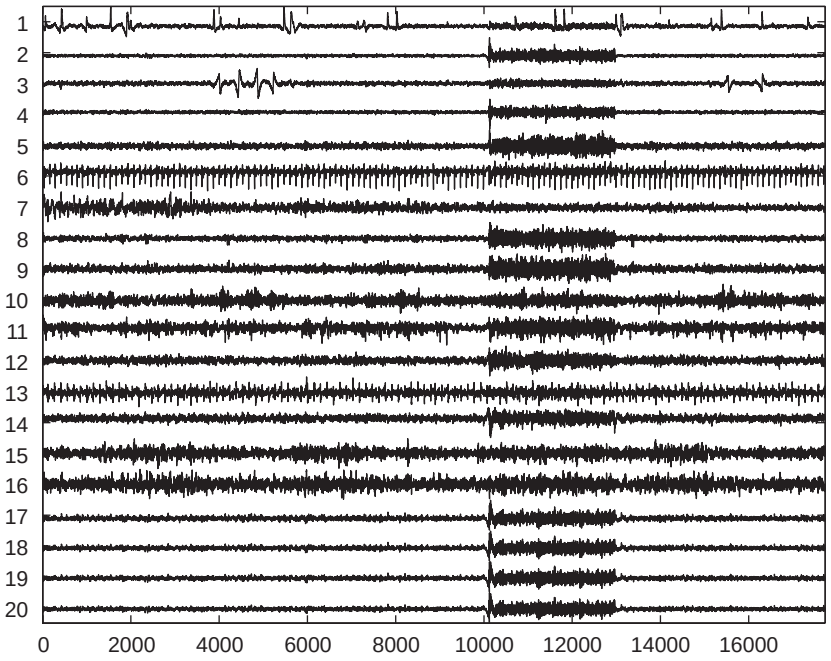

Fig. 11. The 20 source signals obtained by the presented FastNA.

beating and 13 shows clearly the artifact originated from the digital watch. In fact, the extraction results are similar to the well-known ICA [32,33]. Thus, the nonlinear temporal autocorrelation is also likely a strong feature of MEG data sets and the proposed FastNA algorithm provides a fast and robust BSS method.

4. Concluding remarks

We consider the temporal dependencies of source signals as a separation principle. We show that the BSS problem can also be solved by maximizing the nonlinear temporal autocorrelation of sources. Thus, the FastNA algorithm, which is based on a nonlinear measure of temporal autocorrelation, provides a new BSS method.

It is worth noting that the FastNA algorithm can separate the square temporal autocorrelation Gaussian signals which cannot be separated by the well-known ICA algorithms using the non-Gaussianity, such as the FastICA [8]. The simulations shown that the two statistical properties are very different. However, the FastNA and ICA both can separate or extract the real-world FECG and MEG signals. The facts show that the two kinds of data have statistical properties with the non-Gaussianity and the nonlinear temporal autocorrelation at the same time.

When sources have the square temporal autocorrelation (no outlier), we demonstrate that the efficient implementation of the method using two nonlinear functions. In fact, any smooth function could be used for BSS, provided only that the function in the theorem is chosen correctly as the statistical properties of original sources (as the well-known fact in ICA). It is also worth noting that our algorithm obtained here converges very fast and does not need choose any learning step sizes, as the well-known FastICA [8,11].

In ICA and BSS, if the signals to be reconstructed satisfy certain properties, an exact form of the function G is not required in order to achieve the desired estimation results. We may therefore optimistically assume that the exact form of the function G is not very important here either, as long as it is qualitatively similar enough. As a special case, assume that we choose $G(u)=u^2$ and the zero time delay, the presented optimization problem which leads to an algorithm, is formally identical to the one encountered when the value of kurtosis is maximized to find the most non-Gaussian directions in ICA. However, the algorithm using $G(u)=u^2$ is sensitive to the outliers as the kurtosis-based FastICA. But the nonlinear function $G(u)=\log(\cosh(u))$ is better than the kurtosis-based function, due to its better analysis property and robustness to outliers [11].

The choice of the estimation objective function for BSS depends, of course, on the data at hand. The nonlinear temporal autocorrelation is likely a strong feature of data sets in certain applications and can be used for pattern recognition problems such as the statistical feature extraction.

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, Advances in Neural Information Processing Systems 8 (1996) 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, Blind source separation by nonstationarity of variance: a cumulant-based approach, IEEE Transactions on Neural Networks 12 (6) (2001) 1471–1474.
[10] A. Hyvärinen, Complexity pursuit: separating interesting components from time-series, Neural Computation 13 (4) (2001) 883–898.
[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 sub-Gaussian and super-Gaussian 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] Z.Y. Liu, K.C. Chiu, L. Xu, One-bit-matching conjecture for independent component analysis, Neural Computation 16 (2004) 383–399.
[17] K. Matsuoka, M. Ohya, M. Kawamoto, A neural net for blind separation of nonstationary signals, Neural Networks 8 (3) (1995) 411–419.
[18] L. Molgedey, H.G. Schuster, Separation of a mixture of independent signals using time delayed correlations, Physical Review Letters 72 (23) (1994) 3634–3637.
[19] B. De Moor (Ed.), Daisy: database for the identification of systems ⟨http://www.esat.kuleuven.ac.be/sista/daisy⟩ (1997).
[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] 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.
[24] Z. Shi, H. Tang, Y. Tang, A fast fixed-point algorithm for complexity pursuit, Neurocomputing 64 (2005) 529–536.
[25] Z. Shi, C. Zhang, Gaussian moments for noisy complexity pursuit, Neurocomputing 69 (7-9) (2006) 917–921.
[26] Z. Shi, C. Zhang, Energy predictability to blind source separation, Electronics Letters 42 (17) (2006) 1006–1008.
[27] Z. Shi, C. Zhang, Semi-blind source extraction for fetal electrocardiogram extraction by combining non-Gaussianity and time-correlation, Neurocomputing 70 (2007) 1574–1581.
[28] Z. Shi, C. Zhang, Nonlinear innovation to blind source separation, Neurocomputing 71 (2007) 406–410.
[29] Z. Shi, Z. Jiang, F. Zhou, A fixed-point algorithm for blind source separation with nonlinear autocorrelation, Journal of Computational and Applied Mathematics (2008), doi:10.1016/j.cam.2008.03.009.
[30] J.V. Stone, Blind source separation using temporal predictability, Neural Computation 13 (2001) 1559–1574.
[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. Jousmaäki, 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. Jousmaäki, 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] M. Zibulevsky, B.A. Pearlmutter, Blind source separation by sparse decomposition in a signal dictionary, Neural Computation 13 (2001) 863–882.

About the Author—ZHENWEI SHI received his Ph.D. degree in Mathematics from Dalian University of Technology, Dalian, China, in 2005. He was a Postdoctoral Researcher in the Department of Automation, Tsinghua University, Beijing, China, from 2005 to 2007. He is currently an Associate Professor in the Image Processing Center, School of Astronautics, Beihang University. His research interests include blind signal processing, image processing, pattern recognition, machine learning and neuroinformatics.

About the Author—CHANGSHUI ZHANG received his B.S. degree in Mathematics from Peking University, Beijing, China, in 1986, and Ph.D. degree from Department of Automation, Tsinghua University, Beijing, China, in 1992. He is currently a Professor in the Department of Automation, Tsinghua University. He is an Associate Editor of the journal Pattern Recognition. His interests include artificial intelligence, image processing, pattern recognition, machine learning and evolutionary computation.