Neurocomputing 72 (2009) 2556–2562

Contents lists available at ScienceDirect

Neurocomputing

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

Blind source extraction based on generalized autocorrelations and complexity pursuit

Hongjuan Zhang [a], Zhenwei Shi [b], Chonghui Guo [c,*]

[a] Department of Mathematics, Shanghai University, Shanghai 200444, PR China
[b] Image Processing Center, School of Astronautics, Beijing University of Aeronautics and Astronautics, Beijing 100083, PR China
[c] Institute of Systems Engineering, Dalian University of Technology, Dalian 116024, PR China

ARTICLE INFO

Article history:
Received 11 April 2008
Received in revised form
20 August 2008
Accepted 27 October 2008
Communicated by S. Choi
Available online 14 November 2008

Keywords:
Blind source separation
Blind source extraction
Independent component analysis
Autocorrelation
Innovation

ABSTRACT

Blind source extraction (BSE) is a special class of blind source separation (BSS) method. Due to its low computation load and fast processing speed, BSE has become one of the promising methods in signal processing and analysis. This paper addresses BSE problem when a desired source signal has temporal structures. Based on the generalized autocorrelations of the desired signal and the non-Gaussianity of its innovations, we develop an objective function. Maximizing this objective function, we present a BSE algorithm and further give its stability analysis in this paper. Simulations on image data and electrocardiogram (ECG) data indicate its better performance and the better property of tolerance to the estimation error of the time delay.

1. Introduction

The problem of blind source separation (BSS) [1–4] forms part of the independent component analysis (ICA), which is an active research field that has attracted great interest in the field of biomedical signal analysis and processing, geophysical data processing, data mining, speech analysis and image recognition, etc. [5–12]. BSS is a powerful statistical technique in blind signal processing, which considers the simultaneous recovery of all the independent components from the observations. However, in practice, extracting all the source signals from a large number of observed sensor signals, for example, a magnetoencephalographic (MEG) measurement which may output hundreds of recordings, can take a long time and in these signals only a very few are desired with given characteristics. Therefore it is necessary to develop an reliable, robust and effective method to obtain only the desired signals that contain useful information. For this purpose, another technique—blind source extraction (BSE)—was proposed, which has become one of the promising methods due to its low computation load and fast processing speed. Many source extraction algorithms [1] can extract a specific signal as the first output, in which some special characteristics of the signals have to be exploited, such as linear predictability or smoothness [1,13], non-Gaussianity [14,15], sparseness [16] and generalized autocorrelations [17], etc.

When the desired source signals are periodic or quasi-periodic, one convenient way to exploit this property is to employ a linear predictor within the BSE structure. Many source extraction algorithms have considered this case. For example, Barros and Cichocki [13] provided a simple algorithm (simplified with "BCBSE" algorithm) which can quickly extract a desired source signal with a specific period. This algorithm in all the cases could extract the desired sources, whether they are colored as long as they are decorrelated and have temporal structures. However, this method only carries out the constrained minimization of the mean squared error, which cannot well describe the probability distribution of the innovations of the signals. In addition, it needs a prior information about the optimal time delay and is very sensitive to the estimation error of the time delay. To overcome these drawbacks, Shi et al. [15] developed a BSE algorithm (simplified with "SemiBSE" algorithm), which is based on the non-Gaussianity and the autocorrelations of the source signals and contains the mean squared error objective function presented by Barros and Cichocki [13]. This method improves the performance of BCBSE algorithm, and its tolerance to large estimated errors of the period makes the desired signal extracted robustly. However, it must be noted that its better tolerance lies the fact that the choice of initial weight is not random but unit vector. When we initialize the weight randomly, SemiBSE algorithm becomes more sensitive to the estimation error of time delay.

* Corresponding author. Tel.: +86 411 84708007.
E-mail address: guochonghui@tshinghua.org.cn (C. Guo).

0925-2312/$ - see front matter Crown Copyright © 2008 Published by Elsevier B.V. All rights reserved.
doi:10.1016/j.neucom.2008.10.013

The alternative approach, which addresses the BSE problem when a desired source signal has linear or nonlinear autocorrelations, was first introduced in literature [17]. Based on the generalized (i.e., linear or nonlinear) autocorrelations of the primary sources, authors proposed a BSE algorithm (called "GABSE" algorithm). It has been shown that this method has good stability and linear convergence speed. Furthermore, the convergence is global except for a zero measure region. This method only assumes the sources are decorrelated with each other and every source has different temporal structures, but dose not necessarily have to be statistically independent. GABSE algorithm has been applied to many cases directly and the performance is satisfying to a certain extent. Whereas, it also suffers some disadvantages, for example, its tolerance to estimated errors of time delay is not very robust.

In this article, we aim to develop an efficient algorithm, which incorporates the generalized autocorrelations of the desired signal and the non-Gaussianity of its innovations, for extraction of the desired signal. Based on these priori special characteristics, we first develop an objective function. Then a gradient BSE algorithm for finding a maximum of this objective function is proposed. Numerical computation and theoretical analysis show that such a combination both improves the performance of existing algorithms and has better robustness to the estimated error of time delay.

This manuscript is organized as follows. The next section describes the objective function and proposes the BSE algorithm, then gives its stability analysis. Section 3 demonstrates the present technique with experiments using image data and electrocardiogram (ECG) data. The final section provides some discussions and conclusions.

2. Proposed algorithm

2.1. Objective function

Denote the observed sensor signals $\mathbf{x}(t)=(x_1(t),\ldots,x_n(t))^{\mathrm{T}}$ described by matrix equation

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

where $\mathbf{A}$ is an $n\times n$ unknown mixing matrix and $\mathbf{s}(t)=(s_1(t),\ldots,s_n(t))^{\mathrm{T}}$ is a vector of unknown temporally correlated sources. We assume that desired source signal s_i has specific temporal structures—linear or nonlinear autocorrelations. And it can be modelled by linear autoregressive model, which has just one predicting term as

$$s_i(t)=b_i s_i(t-\tau)+\delta_i(t), \tag{2}$$

where $\delta_i(t)$ is a zero-mean, independent identically distributed (*i.i.d.*) time series called innovation and τ is delay in time.

Because we want to extract only a desired source signal, for this purpose we design a single neural processing unit described as

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

$$y(t-\tau)=\mathbf{w}^{\mathrm{T}}\mathbf{x}(t-\tau), \tag{4}$$

$$\hat{y}(t)=y(t)-by(t-\tau), \tag{5}$$

where $y(t)$ and $y(t-\tau)$ are the extracted signals at time t and $(t-\tau)$, respectively, $\mathbf{w}=(w_1,\ldots,w_n)^{\mathrm{T}}$ is the weight vector, $\hat{y}(t)$ represents an innovation and b is a coefficient.

When extracting a desired signal, we are interested in its temporal characteristics (e.g. autocorrelation) of the interesting signal and the probability distribution of its innovations (e.g. non-Gaussianity). Here, assume that the measured sensor signals $\mathbf{x}$ have already been followed by an $n\times n$ whitening filter $\mathbf{V}$ such that the components of $\tilde{\mathbf{x}}(t)=\mathbf{V}\mathbf{x}(t)$ are unit variance and uncorrelated. Then, the extraction of desired signal can be formulated as the following constrained problem by maximizing the convex combination between generalized autocorrelations of a desired signal and the non-Gaussianity of its innovations

$$\max_{\|\mathbf{w}\|=1}\Psi(\mathbf{w})=\lambda E\{G(\tilde{y}(t))G(\tilde{y}(t-\tau))\}-(1-\lambda)E\{F(\tilde{y}(t)-b\tilde{y}(t-\tau))\}, \tag{6}$$

where $\tilde{y}(t)=\mathbf{w}^{\mathrm{T}}\tilde{\mathbf{x}}(t)$, $\tilde{y}(t-\tau)=\mathbf{w}^{\mathrm{T}}\tilde{\mathbf{x}}(t-\tau)$. The coefficient λ is a scalar parameter between 0 and 1 to control the balance between generalized autocorrelation of the desired signal and the non-Gaussianity of its innovations. For instance, only the generalized autocorrelations are considered in the extraction for $\lambda=1$, while only the probability distributions of innovations are optimized for $\lambda=0$. Generally, the parameter is $0<\lambda<1$, depending on the requirement of extraction. The function G is a differentiable function which measures the autocorrelation degree of the desired signal. Examples of choices are $G(u)=u$, $G(u)=u^2$ and $G(u)=\log\cosh(u)$. F is a differentiable function too, which should be determined by the probability distribution of innovations [4,18,19].

2.2. Learning algorithm

To find the maxima of the objective function (6), we can use a simple gradient ascent. The gradient of $\Psi(\mathbf{w})$ with respect to $\mathbf{w}$ can be obtained straightforwardly as

$$\begin{aligned}\frac{\partial\Psi(\mathbf{w})}{\partial\mathbf{w}}&=\lambda E\{g(\tilde{y}(t))G(\tilde{y}(t-\tau))\tilde{\mathbf{x}}(t)+G(\tilde{y}(t))g(\tilde{y}(t-\tau))\tilde{\mathbf{x}}(t-\tau)\}\\&\quad-(1-\lambda)E\{(\tilde{\mathbf{x}}(t)-b\tilde{\mathbf{x}}(t-\tau))f(\tilde{y}(t)-b\tilde{y}(t-\tau))\},\end{aligned} \tag{7}$$

where g and f are the derivatives of the function G and F, respectively. Notice that f should be chosen according to the probability distribution of the innovation. If it is super-Gaussian, $f(u)=\mathrm{sign}(u)$ is suitable, which can be approximated by a smoother function $f(u)=\tanh(au)$, where $a\geqslant 1$. For sub-Gaussian innovation, one can choose one of $f(u)=u-\tanh(u)$ or $f(u)=u^3$. Note that the coefficient b can be learned by the gradient of F.

This means that we should have

$$\begin{aligned}\mathbf{w}&\leftarrow\mathbf{w}+\mu_{\mathbf{w}}(\lambda E\{g(\tilde{y}(t))G(\tilde{y}(t-\tau))\tilde{\mathbf{x}}(t)+G(\tilde{y}(t))g(\tilde{y}(t-\tau))\tilde{\mathbf{x}}(t-\tau)\}\\&\quad-(1-\lambda)E\{(\tilde{\mathbf{x}}(t)-b\tilde{\mathbf{x}}(t-\tau))f(\tilde{y}(t)-b\tilde{y}(t-\tau))\}),\\b&\leftarrow b+\mu_b((1-\lambda)E\{\tilde{y}(t-\tau)f(\tilde{y}(t)-b\tilde{y}(t-\tau))\}).\end{aligned} \tag{8}$$

2.3. Stability analysis

In this section, we provide the stability conditions of the proposed BSE algorithm and its proof. For simplicity of the equations the time index t is omitted in the following, i.e., $s_i=s_i(t)$, $s_{i\tau}=s_i(t-\tau)$, $\delta_i=s_i(t)-b_i s_i(t-\tau)$, and we assume that $E\{s_i^2\}=E\{s_{i\tau}^2\}=1(i=1,\ldots,n)$.

Theorem 1. *Assume that the input data follow the model* (1) *with whitened data* $\tilde{\mathbf{x}}=\mathbf{V}\mathbf{A}\mathbf{s}$, *where* $\mathbf{V}$ *is the whitening matrix and* $\Psi(\mathbf{w})$ *is a sufficiently smooth even function. Furthermore, assume that* $\{s_i,s_{i\tau}\}$ *and* $\{s_j,s_{j\tau}\}$ $(\forall j\neq i)$ *are mutually independent, the innovations* δ_i $(i=1,\ldots,n)$ *are zero mean, and mutually independent. Then, the local maxima of* $\Psi(\mathbf{w})$ *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 *satisfies*

$$\begin{aligned}&\lambda E\{g'(s_i)G(s_{i\tau})+g'(s_{i\tau})G(s_i)+2s_js_{j\tau}g(s_i)g(s_{i\tau})-s_ig(s_i)G(s_{i\tau})\\&\quad-s_{i\tau}G(s_i)g(s_{i\tau})\}+(1-\lambda)E\{\delta_i f(\delta_i)-\delta_j f'(\delta_i)\}<0,\quad\forall j\neq i,\end{aligned} \tag{9}$$

where g and f are the derivatives of G and F, g′ and f′ is the derivative of g and f, respectively, and λ and b_i, $i=1,\ldots,n$ are the known constants.

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

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

which means the matrix **VA** is orthogonal. Make the orthogonal change of coordinates $\mathbf{q} = \mathbf{A}^{\mathrm{T}}\mathbf{V}^{\mathrm{T}}\mathbf{w}$, therefore $\mathbf{w}^{\mathrm{T}}\tilde{\mathbf{x}} = \mathbf{q}^{\mathrm{T}}\mathbf{s}$. Then we can calculate the gradient as

$$\frac{\partial\Psi(\mathbf{q})}{\partial\mathbf{q}} = \lambda E\{\mathbf{s}g(\mathbf{q}^{\mathrm{T}}\mathbf{s})G(\mathbf{q}^{\mathrm{T}}\mathbf{s}_\tau) + \mathbf{s}_\tau G(\mathbf{q}^{\mathrm{T}}\mathbf{s})g(\mathbf{q}^{\mathrm{T}}\mathbf{s}_\tau)\} - (1-\lambda)E\{(\mathbf{s}-\mathbf{b}\cdot\mathbf{s}_\tau) \times f(\mathbf{q}^{\mathrm{T}}\mathbf{s}-\mathbf{b}\cdot\mathbf{q}^{\mathrm{T}}\mathbf{s}_\tau)\} \tag{11}$$

and the Hessian as

$$\begin{aligned}\frac{\partial^2\Psi(\mathbf{q})}{\partial\mathbf{q}^2} = {}& \lambda E\{\mathbf{s}\mathbf{s}^{\mathrm{T}}g'(\mathbf{q}^{\mathrm{T}}\mathbf{s})G(\mathbf{q}^{\mathrm{T}}\mathbf{s}_\tau) + \mathbf{s}\mathbf{s}_\tau^{\mathrm{T}}g(\mathbf{q}^{\mathrm{T}}\mathbf{s})g(\mathbf{q}^{\mathrm{T}}\mathbf{s}_\tau) + \mathbf{s}_\tau^{\mathrm{T}}\mathbf{s}g(\mathbf{q}^{\mathrm{T}}\mathbf{s})g(\mathbf{q}^{\mathrm{T}}\mathbf{s}_\tau) \\ & + \mathbf{s}_\tau\mathbf{s}_\tau^{\mathrm{T}}g'(\mathbf{q}^{\mathrm{T}}\mathbf{s}_\tau)G(\mathbf{q}^{\mathrm{T}}\mathbf{s})\} - (1-\lambda)E\{(\mathbf{s}-\mathbf{b}\cdot\mathbf{s}_\tau)(\mathbf{s}-\mathbf{b}\cdot\mathbf{s}_\tau)^{\mathrm{T}} \\ & \times f'(\mathbf{q}^{\mathrm{T}}\mathbf{s}-\mathbf{b}\cdot\mathbf{q}^{\mathrm{T}}\mathbf{s}_\tau)\},\end{aligned} \tag{12}$$

where $\mathbf{b} = (b_1,\ldots,b_n)^{\mathrm{T}}$. Without loss of generality, it is enough to analyze the stability of the point $\mathbf{q} = \mathbf{e}_1$, where $\mathbf{e}_1 = (1,0,\ldots,0)^{\mathrm{T}}$. Evaluating the gradient and the Hessian at the point $\mathbf{q} = \mathbf{e}_1$

$$\frac{\partial\Psi(\mathbf{e}_1)}{\partial\mathbf{q}} = \lambda E\{s_1 g(s_1)G(s_{1\tau}) + s_{1\tau}G(s_1)g(s_{1\tau})\} - (1-\lambda)E\{\delta_1 f(\delta_1)\}\mathbf{e}_1 \tag{13}$$

and

$$\begin{aligned}\frac{\partial^2\Psi(\mathbf{e}_1)}{\partial\mathbf{q}^2} = {}& \mathrm{diag}(\lambda E\{s_1^2 g'(s_1)G(s_{1\tau}) + 2s_1 s_{1\tau}g(s_1)g(s_{1\tau}) + s_{1\tau}^2 g'(s_{1\tau})G(s_1)\} \\ & - (1-\lambda)E\{\delta_1^2 f'(\delta_1)\}, \lambda E\{g'(s_1)G(s_{1\tau}) + 2s_2 s_{2\tau}g(s_1)g(s_{1\tau}) \\ & + g'(s_{1\tau})G(s_1)\} - (1-\lambda)E\{\delta_2^2 f'(\delta_1)\},\ldots,\lambda E\{g'(s_1)G(s_{1\tau}) \\ & + 2s_n s_{n\tau}g(s_1)g(s_{1\tau}) + g'(s_{1\tau})G(s_1)\} - (1-\lambda)E\{\delta_n^2 f'(\delta_1)\}).\end{aligned} \tag{14}$$

Making a small perturbation $\varepsilon = (\varepsilon_1,\varepsilon_2,\ldots,\varepsilon_n)^{\mathrm{T}}$, we obtain

$$\begin{aligned}\Psi(\mathbf{e}_1+\varepsilon) &= \Psi(\mathbf{e}_1) + \varepsilon^{\mathrm{T}}\frac{\partial\Psi(\mathbf{e}_1)}{\partial\mathbf{q}} + \frac{1}{2}\varepsilon^{\mathrm{T}}\frac{\partial\Psi^2(\mathbf{e}_1)}{\partial\mathbf{q}^2}\varepsilon + o(\|\varepsilon\|^2) \\ &= \Psi(\mathbf{e}_1) + [\lambda E\{s_1 g(s_1)G(s_{1\tau}) + s_{1\tau}G(s_1)g(s_{1\tau})\} \\ &\quad - (1-\lambda)E\{\delta_1 f(\delta_1)\}]\varepsilon_1 + \frac{1}{2}[\lambda E\{s_1^2 g'(s_1)G(s_{1\tau}) \\ &\quad + 2s_1 s_{1\tau}g(s_1)g(s_{1\tau}) + s_{1\tau}^2 g'(s_{1\tau})G(s_1)\} \\ &\quad - (1-\lambda)E\{\delta_1^2 f'(\delta_1)\}]\varepsilon_1^2 + \frac{1}{2}\sum_{j>1}[\lambda E\{g'(s_1)G(s_{1\tau}) \\ &\quad + 2s_j s_{j\tau}g(s_1)g(s_{1\tau}) + g'(s_{1\tau})G(s_1)\} \\ &\quad - (1-\lambda)E\{\delta_j^2 f'(\delta_1)\}]\varepsilon_j^2 + o(\|\varepsilon\|^2).\end{aligned} \tag{15}$$

Due to the constraint $\|\mathbf{w}\| = 1$ and the matrix **VA** being orthogonal, we have $\|\mathbf{q}\| = \|\mathbf{A}^{\mathrm{T}}\mathbf{V}^{\mathrm{T}}\mathbf{w}\| = 1$. Thus, we get $\varepsilon_1 = \sqrt{1-\varepsilon_2^2-\varepsilon_3^2-\cdots-\varepsilon_n^2} - 1$. Due to the fact that $\sqrt{1-\gamma} = 1-\gamma/2+o(\gamma)$, the term of order ε_1^2 in (15) is $o(\|\varepsilon\|^2)$, i.e., of higher order, and can be neglected. Using the aforementioned first-order approximation for ε_1 we obtain $\varepsilon_1 = -\sum_{i>1}\varepsilon_i^2/2 + o(\|\varepsilon\|^2)$, which finally gives

$$\begin{aligned}\Psi(\mathbf{e}_1+\varepsilon) = {}& \Psi(\mathbf{e}_1) + \frac{1}{2}\sum_{j>1}[\lambda E\{g'(s_1)G(s_{1\tau}) + g'(s_{1\tau})G(s_1) \\ & + 2s_j s_{j\tau}g(s_1)g(s_{1\tau}) - s_1 g(s_1)G(s_{1\tau}) - s_{1\tau}G(s_1)g(s_{1\tau})\} \\ & + (1-\lambda)E\{\delta_1 f(\delta_1) - \delta_j^2 f'(\delta_1)\}]\varepsilon_j^2 + o(\|\varepsilon\|^2),\end{aligned} \tag{16}$$

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

Remark 1. Assume that the parameter $\lambda = 1$ is chosen, the conclusion of Theorem 1 is the same as that of the GABSE [17], i.e.

$$\begin{aligned}&E\{g'(s_i)G(s_{i\tau}) + g'(s_{i\tau})G(s_i) + 2s_j s_{j\tau}g(s_i)g(s_{i\tau}) - s_i g(s_i)G(s_{i\tau}) \\ &\quad - s_{i\tau}g(s_{i\tau})G(s_i)\} < 0, \quad \forall j\neq i.\end{aligned} \tag{17}$$

Remark 2. Assume that the parameter $\lambda = 0$ and $E\{\delta_i^2\} = 1$, the local maxima is obtained such that

$$E\{\delta_i f(\delta_i) - f'(\delta_i)\} < 0, \tag{18}$$

which is identical to that of [15].

3. Simulations

In order to verify the efficiency of our algorithm, we made many simulations on image data, artificial ECG data and real-world ECG data [20]. Moreover, we compared the proposed method with many existing techniques—BCBSE algorithm, SemiBSE algorithm, GABSE algorithm and one-unit FastICA algorithm [21]. In these comparisons, the performance of algorithms to estimate the desired signal is measured by performance index (PI), which is defined as follows:

$$\mathrm{PI} = \sum_{j=1}^{n}\frac{|p_j|}{\max_k|p_k|} - 1, \quad k = 1,\ldots,n, \tag{19}$$

where p_j denotes the j element of the global vector $\mathbf{p} = \mathbf{w}^{\mathrm{T}}\mathbf{VA}$. PI is zero when the desired signal is perfectly extracted. Besides, the accuracy of the extracted signal compared to the source is expressed using the signal-to-noise ratio (SNR) in dB given by

$$\mathrm{SNR} = 10\log_{10}(s^2/\mathrm{MSE}), \tag{20}$$

where s^2 denotes the variance of the source signal, MSE denotes the mean square error between the original signal and the extracted signal. The higher SNR is, the better performance is.

3.1. Experiments on image data

Two 128 × 128 images and one *i.i.d.* Gaussian noise were used in the first simulation. From the top to down, Fig. 1 gives the original images sources 1–3, the mixtures 1–3.

For finding the appropriate delay τ, we calculated the autocorrelation $\zeta(k) = E\{x_j(t)x_j(t-k)\}$ $(j=1,2,3)$ of the sensor

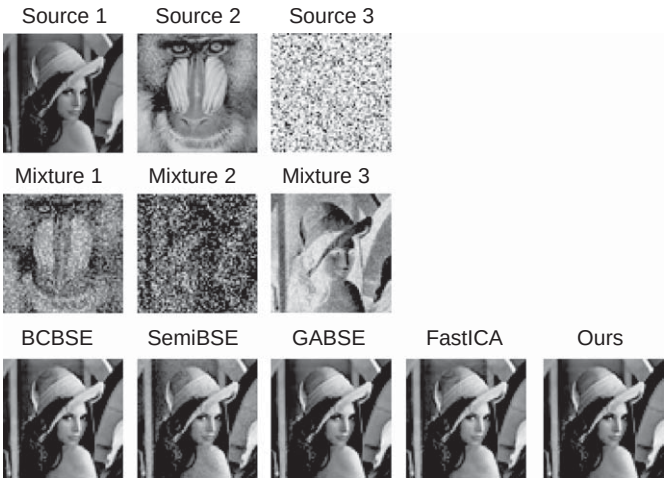

Fig. 1. Simulation results for mixture of two images and one *i.i.d.* Gaussian signal. From top to bottom, the original images sources 1–3, the mixtures 1–3, and the extractions by BCBSE, SemiBSE, GABSE, FastICA and our algorithm.

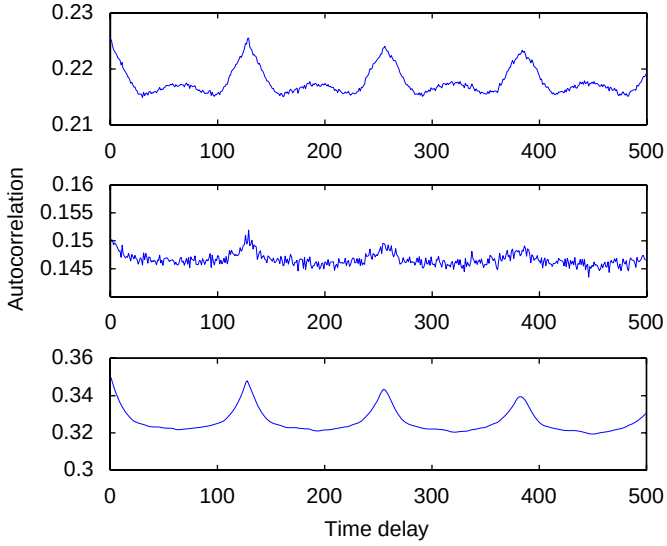

Fig. 2. Autocorrelation functions of the mixtures 1–3 of Fig. 1.

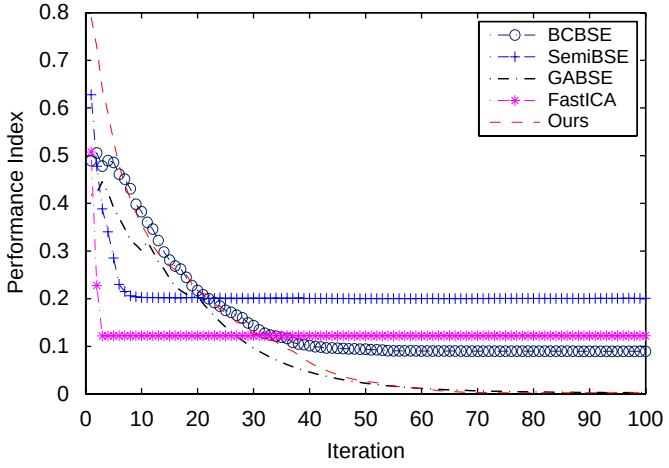

Fig. 3. Comparisons of median PIs over 100 independent runs for five algorithms at $\tau = 128$ on image data.

signals as a function of the time delay k and found the optimal time delay τ at which the autocorrelation function has a peak [13,17]. The corresponding autocorrelation functions of three mixtures are shown in Fig. 2. One can see that the autocorrelation functions present several peaks and the peaks are similar for each autocorrelation function, thus, we chose the first optimal time delay $\tau = 128$.

The extracted images by BCBSE, SemiBSE, GABSE FastICA and our algorithm are presented in the bottom of Fig. 1. The SNR values of them are 21.2819 dB (BCBSE), 14.4324 dB (SemiBSE), 51.3892 dB (GABSE), 18.9753 dB (FastICA) and 52.7096 dB (our algorithm), respectively. The superiority of our algorithm is straightforward. Moreover, this experiment is independently repeated 100 times and the averaged SNRs are 21.1807 dB (BCBSE), 13.8797 dB (SemiBSE), 50.4471 dB (GABSE), 5.4115 dB (FastICA) and 54.1878 dB (our algorithm), respectively. It is noting that for our algorithm the highest SNR index among 100 experiments reaches 67.2184 dB, however, GABSE algorithm is only 51.8084 dB.

Note that the nonlinear functions in SemiBSE and FastICA algorithm are chosen as $\log\cosh(u)$ and linear function is used in GABSE algorithm. In our algorithm, $G(u)$ and $F(u)$ are selected as linear function and $\log\cosh(u)$, respectively. The learning rates of SemiBSE algorithm are set to be 0.5 ($\mu_{\mathbf{w}}$) and 0.0005 (μ_b), and those of our algorithm are 0.2 ($\mu_{\mathbf{w}}$) and 0.00001 (μ_b).

In our algorithm, the parameter λ aims to balance generalized autocorrelation of the desired signal and the non-Gaussianity of its innovations. By adjusting this parameter, we obtain an optimal solution with the objective function 0.6728 for $\lambda = 0.74$. Such a result is actually robust with the parameter perturbations, e.g. we can obtain the same corresponding result in other parameters, such as $\lambda = 0.5$ and 0.9. In addition, it must be noting that FastICA algorithm is confident of *Source*1 as the first recovery, for which the cost function values of three source signals are computed as {0.2932, 0.2432, 0.0547}, respectively, and *Source*1 corresponds to the largest one, where the cost function used here is the one in literatures [4,21].

For the comparison of performance index (19) at $\tau = 128$, we test five algorithms in image experiment. The performance is estimated as the median PI values of 100 independent trials. At every trial, five algorithms are run with 100 iterations, which seem to be always enough for convergence. Here **A** and **w** are initialized randomly. The results are depicted in Fig. 3. Obviously, the proposed algorithm performs more efficiently than other algorithms.

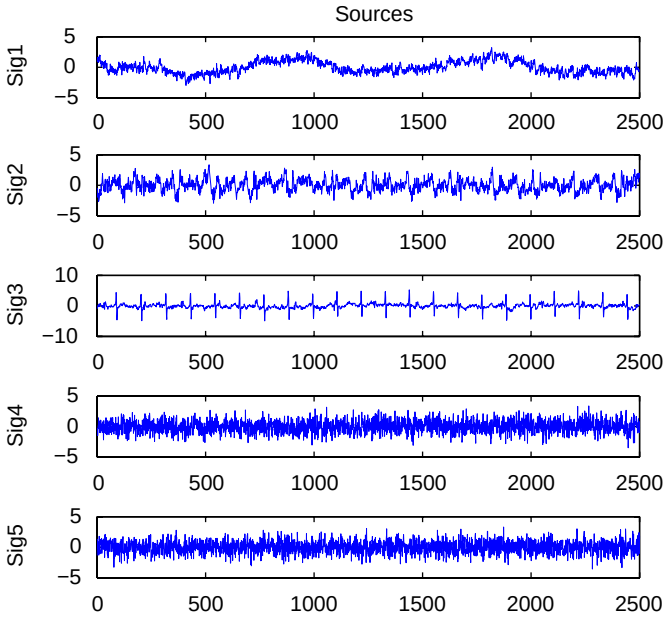

Fig. 4. Five artificial ECG signals. (Sig1) Breathing artifact. (Sig2) Electrode artifact. (Sig3) FECG. (Sig4–Sig5) Two *i.i.d.* Gaussian signals.

3.2. Experiments on ECG data

One important real-world application of BSE is the extraction of fetal ECG (FECG); the aim is to obtain the clear FECG as the first extracted signal [13]. The FECG contains valuable information about the health and condition of the fetus. However, it is always corrupted by various kinds of noise, such as the maternal ECG (MECG) with extreme high amplitude, respiration and stomach activity, thermal noise, noise from electrode-skin contact, etc. In the following experiments, we compare our algorithms and many existing algorithms by artificial ECG data and real-world ECG data.

3.2.1. Experiments on artificial ECG data

We adopted five zero-mean and unit-variance source signals (2500 samples), which are shown in Fig. 4. From the top to down,

they are, a breathing artifact, electrode artifact, an FECG ($\tau = 112$), and two *i.i.d.* Gaussian signals. The observed signals are generated by a 5×5 random mixing matrix. We ran BCBSE, SemiBSE, GABSE, FastICA and our algorithm, respectively. The extracted FECG signals are shown in Fig. 5.

The accuracies of them are 3.2434 dB (BCBSE), 10.6897 dB (SemiBSE), 24.1533 dB (GABSE), 27.6828 dB (FastICA) and 28.5162 dB (our algorithm). The extracted FECG signal by our algorithm is the best, and those extracted by FastICA and GABSE

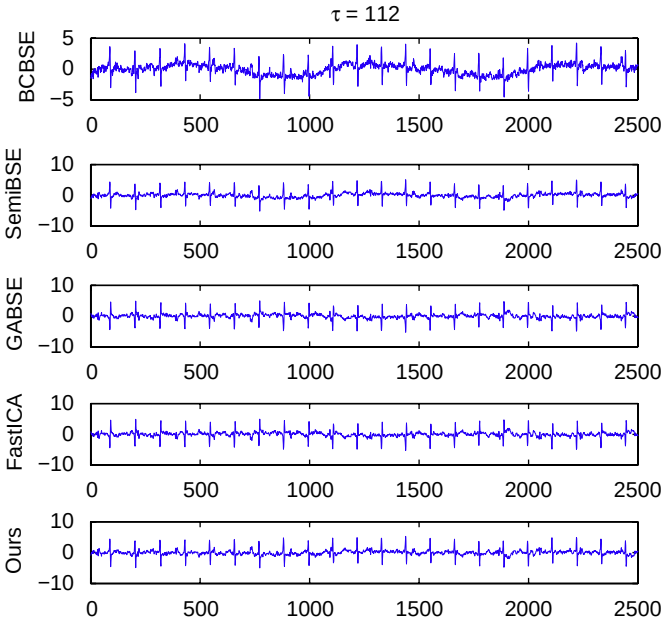

Fig. 5. Extractions for artificial ECG signals at $\tau = 112$. From top to bottom, the extracted FECGs by BCBSE, SemiBSE, GABSE, FastICA and our algorithm.

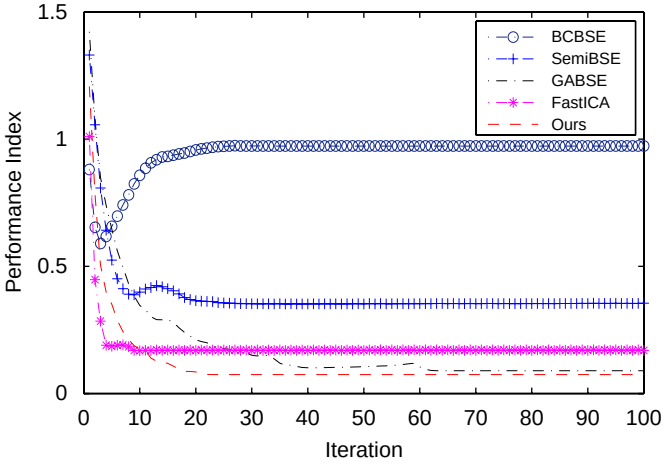

Fig. 6. Comparisons of average PIs over 100 independent runs for five algorithms at $\tau = 112$ on artificial ECG data.

algorithm are also satisfying. The extractions by SemiBSE and BCBSE algorithm are the worst, which are mixed by some or lots of respiration noise. Note that FastICA algorithm must take the third source signal as the first output because the value of its cost function is the largest one. In fact, the cost function values of above five sources are {0.0074, 0.0073, 0.0834, 0.0122, 0.0167}, respectively. And the time delay τ of other algorithms is all chosen as 112. In SemiBSE, GABSE and FastICA algorithm, the nonlinear functions are all chosen as $\log\cosh(\cdot)$. $G(u) = u^2$, $F(u) = \log\cosh(u)$ are used in our algorithm. The learning rates $\mu_{\mathbf{w}}$ and μ_b of our algorithm are all set to be 1. The parameters of other algorithms are the same as those of above-mentioned image extraction experiment. Using our method, we obtain an optimal solution with the objective function 0.9329 for $\lambda = 0.15$. Note that Fig. 6 shows the averaged PIs over 100 independent trials. Obviously, the proposed algorithm has the lowest PI values than other algorithms.

It is mentioning about that our algorithm has the perfect property of robustness. Even if some estimated errors of the time delay were introduced, our algorithm also worked well, which is important in practice. When τ was varied from 110 to 118, we applied four algorithms on the mixtures of artificial ECG data. Table 1 shows the comparison results. Notice that the SNR index for given τ was estimated by the average SNR value of 100 independent trials. Meanwhile, BCBSE, SemiBSE and GABSE our algorithm were only used in the comparisons, not including FastICA algorithm, for which only these three algorithms were influenced by the time delay τ. In each of experiments, the mixing matrix $\mathbf{A}$ and the weight vectors $\mathbf{w}$ were generated randomly. It can be shown that on the whole, our algorithm is better than GABSE algorithm, and superior to BCBSE and SemiBSE algorithm.

3.2.2. *Experiments on real-world ECG data*

In the following, we have performed experiments on real-world ECG data which is distributed by De Moor [20]. These data are a famous ECG measured from a pregnant woman (in Fig. 7). One can see the heart beating of both the mother (stronger and slower) and the fetus (weaker and faster). Note that the fetal influence is stronger in the first channel of Fig. 7. The ECG measurements are recorded over 10 s and sampled at 250 Hz (although in De Moor's homepage he claims the sampling frequency is 500 Hz, Barros et al. [13] assure it is 250 Hz). By using prior information about FECG frequency, we may estimate the optimal time delay $\tau = 112$. The choice of nonlinear functions, the learning rates and other parameters of five algorithms are the same as the foregoing artificial ECG simulations. Fig. 8 provides the extracted FECGs by BCBSE, SemiBSE, GABSE, FastICA and our algorithm, respectively. It shows the desired FECGs are well extracted by all algorithms except FastICA algorithm, which corresponds to clearer MECG signal. The reason may lie in the fact that the pure MECG is a higher super-Gaussian signal than FECG source signal. Therefore MECG signal should be obtained as the first extraction of FastICA algorithm. Moreover, we can see that our algorithm and GABSE algorithm are able to extract clearer

Table 1
Comparison of average SNRs (dB) of the extracted FECGs by four algorithms at different time delays for artificial ECG data.

τ	110	111	112	113	114	115	116	117	118
BCBSE	−3.047	0.923	3.243	−2.575	−3.072	−3.052	−3.05	−3.035	−3.019
SemiBSE	−0.557	7.752	10.731	12.287	10.367	9.580	0.321	−1.4005	−3.001
GABSE	27.943	25.928	24.153	21.860	25.419	20.877	11.62	7.214	−2.38
Ours	26.037	31.117	27.311	26.025	24.738	28.93	13.231	11.01	14.944

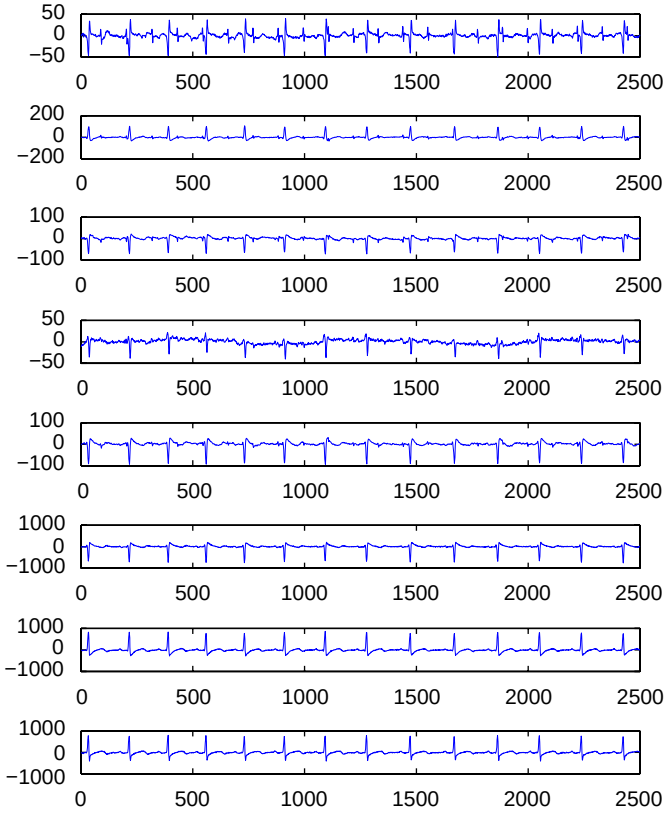

Fig. 7. The 8-channel of ECG recording obtained from a pregnant woman.

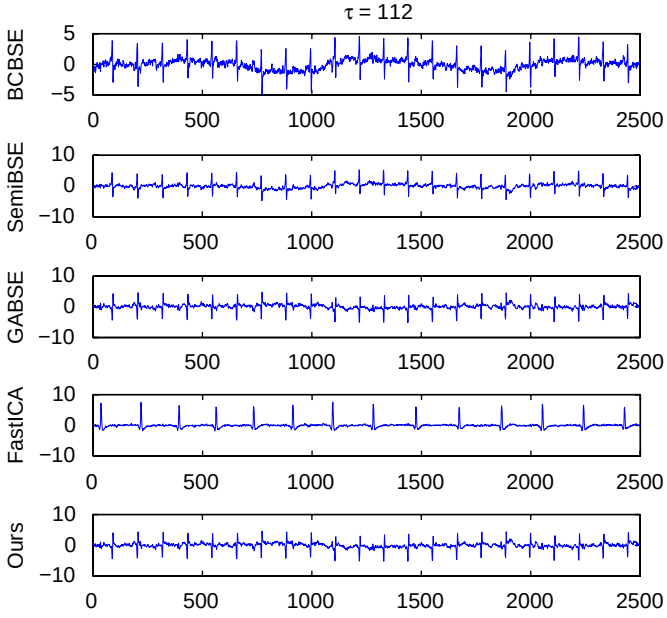

Fig. 8. Comparisons of extracted FECGs by BCBSE, SemiBSE, GABSE, FastICA and our algorithm at the optimal time delay $\tau = 112$ for real-world ECG data.

FECGs, however, extractions by BCBSE and SemiBSE algorithm include many respiration noise. It should be noticed that, since the mixing matrix **A** and the pure FECG signal are not available, the performance index, such as SNR and PI cannot be computed as above. But we can perceive distinctly the quality of extracted FECG through experience.

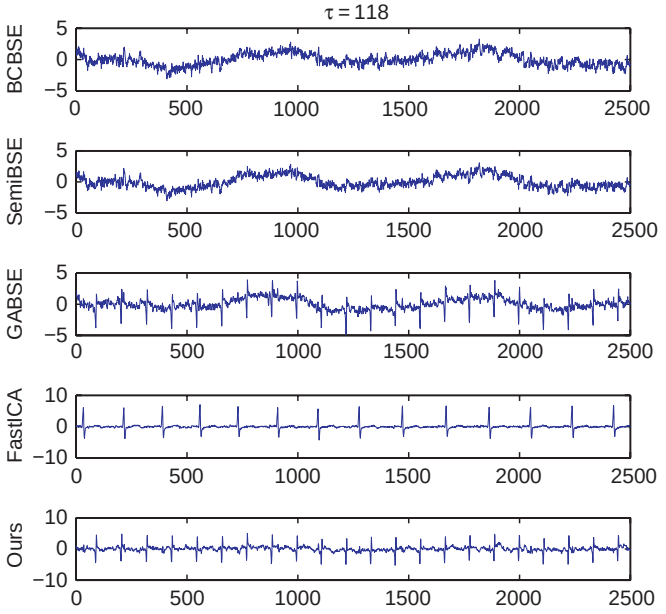

Fig. 9. Comparisons of extracted FECGs by BCBSE, SemiBSE, GABSE, FastICA and our algorithm at $\tau = 118$ for real-world ECG data.

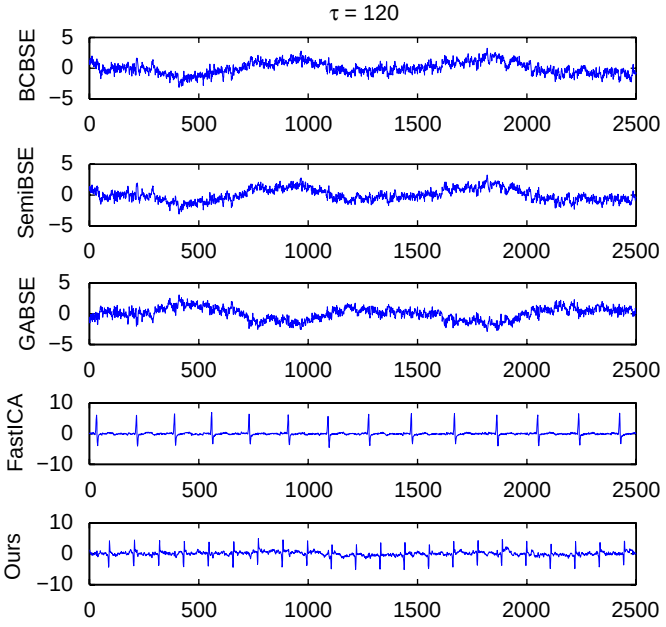

Fig. 10. Comparisons of extracted FECGs by BCBSE, SemiBSE, GABSE, FastICA and our algorithm at $\tau = 120$ for real-world ECG data.

From the experiments of artificial ECG data, we know that the proposed algorithm has better robustness to the estimation error of the time delay. In order to confirm this conclusion further, we make extraction experiments on real-world ECG data when $\tau =$ 118 and 120. The extractions at $\tau = 118, 120$ are plotted in Figs. 9 and 10, respectively. It can be seen that BCBSE and SemiBSE algorithm seem to be incapable of extracting FECG signal at these two time delays. And the extractions of FastICA algorithm are always the clearer MECG signals at these time delays. As for GABSE algorithm, its performance has gradually deteriorated until complete failure. On the contrary, the satisfactory FECG signals are extracted by our algorithm at $\tau = 118$ and 120.

4. Discussions and conclusions

Due to the low computation load and fast processing speed, BSE has become one of the promising method in the field of neural networks, especially unsupervised learning, and more generally in advanced statistics and signal processing. Several approaches for extracting desired signals have been developed recently, such as SemiBSE applied to FECG extraction and GABSE applied to image and FECG extraction. These methods have their advantages or features mainly to the signals of interest with temporal structures. For example, SemiBSE evaluates FECG signal by adopting the non-Gaussianity and the autocorrelations of the source signals, while GABSE analyzes the generalized autocorrelations of the primary sources by a fixed-point algorithm. In contrast, we proposed an effective algorithm which can extract the desired signal as the first output in a more accurate manner.

Specifically, we develop a novel formulation as well as an algorithm based on gradient method for extraction, which aims to utilize the generalized autocorrelations of the desired signal and the non-Gaussianity of its innovations. In contrast to SemiBSE and GABSE algorithm, our approach can handle the extraction problem with higher quality and more robust. To evaluate the autocorrelations of desired signal and exploit the probability distribution of its innovations, we propose an objective function with a convex combination form. Depending on the parameter λ which balances generalized autocorrelation of the desired signal and the non-Gaussianity of its innovations. A large λ emphasizes on the autocorrelations of desired signal. On the other hand, a small λ emphasizes on the non-Gaussianity of its innovations. Maximizing this objective function, we present a BSE algorithm and further give its stability analysis in this paper. As demonstrated in this article, the proposed method can be applied to various types of data, such as image and ECG data, which show its better performance. Note that the ability of robustness to the estimation error of the time delay is very useful in practice, which makes the extraction robustly even if there exist some estimated errors of the time delay.

Acknowledgments

This work is supported by Natural Science Foundation of China under Grant nos. 10571018, 60605002, 70431001, 70871015. The authors would like to thank the referees and the editorial board for their insightful comments and suggestions.

References

[1] A. Cichocki, S. Amari, Adaptive Blind Signal and Image Processing: Learning Algorithms and Applications, Wiley, New York, 2002.
[2] P. Comon, Independent component analysis: a new concept?, Signal Process. 36 (1994) 287–314.
[3] S.A. Cruces-Alvarez, A. Cichocki, S. Amari, From blind signal extraction to blind instantaneous signal separation: criteria, algorithm, and stability, IEEE Trans. Neural Networks 15 (4) (2004) 859–873.
[4] A. Hyvärinen, J. Karhunen, E. Oja, Independent Component Analysis, Wiley, New York, 2001.
[5] S. Amari, A. Cichocki, Adaptive blind signal processing—neural network approaches, Proc. IEEE 86 (10) (1998) 2016–2048.
[6] K. Anand, G. Mathew, V. Reddy, Blind separation of multiple co-channel BPSK signals arriving at an antenna array, IEEE Signal Process. Lett. 2 (9) (1995) 176–178.
[7] S. Boudet, L. Peyrodie, P. Gallois, C. Vasseur, Filtering by optimal projection and application to automatic artifact removal from EEG, Signal Process. 87 (8) (2007) 1978–1992.
[8] E. Chaumette, P. Comon, D. Muller, ICA-based technique for radiating sources estimation: application to airport surveillance, IEE Proc. F 140 (6) (1993) 395–401.
[9] A. Cichocki, T. Rutkowski, A.K. Barros, S.H. Oh, A blind extraction of temporally correlated but statistically dependent acoustic signals, in: Neural Networks for Signal Processing X: Proceedings of the 2000 IEEE Signal Processing Society Workshop (NNSP2000), IEEE Signal Processing Society, Sydney, Australia, 2000, pp. 455–464.
[10] K.F. Côco, E.O.T. Salles, M. Sarcinelli-Filho, Topographic independent component analysis based on fractal theory and morphology applied to texture segmentation, Signal Process. 87 (8) (2007) 1966–1977.
[11] K.E. Hild, H.T. Attias, S. Comani, S.S. Nagarajan, Fetal cardiac signal extraction from magnetocardiographic data using a probabilistic algorithm, Signal Process. 87 (8) (2007) 1993–2004.
[12] L. De Lathauwer, B. De Moor, J. Vandewalle, Fetal electrocardiogram extraction by source subspace separation, in: Proceedings of the HOS'95, Aiguablava, Spain, 1995, pp. 134–138.
[13] A.K. Barros, A. Cichocki, Extraction of specific signals with temporal structure, Neural Comput. 13 (2001) 1995–2003.
[14] A. Hyvärinen, E. Oja, A fast fixed-point algorithm for independent component analysis, Neural Comput. 9 (7) (1997) 1483–1492.
[15] Z. Shi, C. Zhang, Semi-blind source extraction for fetal electrocardiogram extraction by combining non-Gaussianity and time-correlation, Neurocomputing 70 (2007) 1574–1581.
[16] M. Zibulevsky, Y.Y. Zeevi, Extraction of a source from multichannel data using sparse decomposition, Neurocomputing 49 (2002) 163–173.
[17] Z. Shi, C. Zhang, Blind source extraction using generalized autocorrelations, IEEE Trans. Neural Networks 18 (5) (2007) 1516–1524.
[18] A. Hyvärinen, Complexity pursuit: separating interesting components from time-series, Neural Comput. 13 (4) (2001) 883–898.
[19] Z. Shi, H. Tang, Y. Tang, A fast fixed-point algorithm for complexity pursuit, Neurocomputing 64 (2005) 529–536.
[20] D. De Moor (Ed.), Daisy: database for the identification of systems, available online at: ⟨http://www.esat.kuleuven.ac.be/sista/daisy⟩.
[21] A. Hyvärinen, Fast and robust fixed-point algorithm for independent component analysis, IEEE Trans. Neural Networks 10 (3) (1999) 626–634.

Hongjuan Zhang received her B.S. degree in Mathematics from Ludong University, China, in 2003, M.S. degree from Department of Applied Mathematics of Dalian University of Technology in 2005. She is now a doctorate student in Department of Applied Mathematics of Dalian University of Technology. Her research interests cover blind signal processing, pattern recognition, and systems optimization.

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, Beijing University of Aeronautics and Astronautics. His research interests include blind signal processing, image processing, pattern recognition, machine learning and neuroinformatics.

Chonghui Guo received the B.S. degree in Mathematics from Liaoning University, China, in 1995, M.S. degree in Operational Research and Control Theory, and Ph.D. degree in Management Science and Engineering from Dalian University of Technology, China, in 2002. He was a postdoctoral research fellow in the Department of Computer Science, Tsinghua University from 2002 to 2004. Now he is an associate professor of the Institute of Systems Engineering, Dalian University of Technology. His current research interests include machine learning, data mining, pattern recognition, and systems optimization.