INTERNATIONAL JOURNAL OF ADAPTIVE CONTROL AND SIGNAL PROCESSING
Int. J. Adapt. Control Signal Process. 2004; **18**:279–298 (DOI: 10.1002/acs.794)

Blind separation of convolutive mixtures of cyclostationary signals

Wenwu Wang*,†, Maria G. Jafari, Saeid Sanei and Jonathon A. Chambers

Centre for Digital Signal Processing Research, Department of Electronic Engineering, King's College London, London WC2R 2LS, U.K.

SUMMARY

An adaptive blind source separation algorithm for the separation of convolutive mixtures of cyclostationary signals is proposed. The algorithm is derived by applying natural gradient iterative learning to a novel cost function which is defined according to the wide sense cyclostationarity of signals and can be deemed as a new member of the family of natural gradient algorithms for convolutive mixtures. A method based on estimating the cycle frequencies required for practical implementation of the proposed algorithm is presented. The efficiency of the algorithm is supported by simulations, which show that the proposed algorithm has improved performance for the separation of convolved cyclostationary signals in terms of convergence speed and waveform similarity measurement, as compared to the conventional natural gradient algorithm for convolutive mixtures. Copyright © 2004 John Wiley & Sons, Ltd.

KEY WORDS: blind source separation (BSS); convolutive mixtures; cyclostationary signals; natural gradient learning; cycle frequency determination

1. INTRODUCTION

In many practical situations such as in the radiocommunications, telemetry, radar, sonar and speech contexts, the sources are non-stationary and very often (quasi)-cyclostationary and the observed signals are usually convolutive mixtures, so that the conventional methods for the standard blind source separation (BSS) problem, in which the mixtures are assumed to be instantaneous and the source signals are assumed to be statistically stationary [1], are no longer appropriate. Increasing interest has therefore been focused on solving the problem of BSS of convolutive mixtures and the main existing strategies can be approximately classified into [2]: (1) carrying out blind separation in the time domain by extending the existing BSS algorithms for the instantaneous case to the convolutive case [3]; (2) transforming totally or partially the

*Correspondence to: Dr. Wenwu Wang, Department of Electronic Engineering, King's College London, Strand, London WC2R 2LS, U.K.
† E-mail: wenwu.wang@kcl.ac.uk

Contract/grant sponsor: Engineering and Physical Sciences Research Council, U.K.

Published online 2 March 2004

Revised 10 March 2003
Accepted 17 October 2003

convolutive BSS problem into multiple instantaneous problems in the frequency domain and separating the instantaneous mixtures in every frequency bin [4, 5]; (3) decomposing the problem rather than learning the possibly huge filters all at once, i.e. the decorrelation approach, subspace method, or subband decomposition approach [6]; (4) exploiting the statistical properties or special structure contained within the source signals to formulate various separation criteria [7].

Addressing the BSS problem for cyclostationary sources is a relatively new approach. Meraim *et al.* [7] proposed to minimize a cost function constructed from the cyclic cross-correlation of recovered sources at various time lags, and they presented iterative update equations following from the natural gradient technique. Ferreol and Chevalier [8] have shown that the current second- or higher-order BSS methods perform poorly if the assumption that the source signals are statistically stationary remains unchanged. Most existing BSS approaches exploiting the cyclostationarity of the sources are either batch algorithms as in Reference [8] or are based on second-order statistics as in Reference [7], and very few of them are concerned with the convolutive mixtures. In this paper, we propose a sequential BSS algorithm for convolutive mixtures using high-order conventional statistics and second-order cyclostationarity of the source signals. We will show that the statistical property of cyclostationary signals can be exploited to increase the separability (or separation quality) of the convolutive mixtures of cyclostationary sources. The ensuing sections are organized as follows. The data model is formulated in Section 2 and the proposed algorithm for instantaneous mixtures and convolutive mixtures is detailed in Section 3 which also includes the analysis of the stability condition of the algorithm. A proposed estimation method for the cycle frequencies which are required for the proposed algorithm is presented in Section 4. The simulation results and conclusions are, respectively, presented in Sections 5 and 6. The derivation of several equations in Section 3 is attached in Section 7 (i.e. the appendix).

2. PROBLEM FORMULATION

Assume that N source signals are recorded by M sensors, where $M \geqslant N$. The output of each sensor is modelled as a weighted sum of convolutions of the source signals corrupted by additive noise. In a compact form, we have

$$\mathbf{x}(k) = \mathbf{H}(z)\mathbf{s}(k) + \mathbf{v}(k) \tag{1}$$

where $\mathbf{s}(k) \in \mathbb{C}^N$ is the source vector, $\mathbf{x}(k) \in \mathbb{C}^N$ is the sensor vector, k is the discrete time index; $\mathbf{H}(z)$ is the z-transform of the mixing matrix with entries $H_{ij}(z) = \sum_{p=0}^{P-1} h_{ijp}z^{-p}$ ($i = 1, \ldots, M$; $j = 1, \ldots, N$), where z^{-1} is the time-shift operator, i.e. $z^{-1}s_j(k) = s_j(k-1)$. For simplicity, we ignore additive noise in the following derivations. The source signal vector $\mathbf{s}(k)$ is modelled as a wide sense cyclostationary signal [9], and its components $s_i(k)$ are assumed to be mutually independent with zero mean.

The aim is to reconstruct the source signals $s_i(k)$ (up to an arbitrary permutation and filtering operation) from only knowledge of the sensor signals $x_i(k)$ without knowing the source signals and the mixing process. Alternatively, we have

$$\mathbf{y}(k) = \mathbf{W}(z,k)\mathbf{x}(k) \tag{2}$$

$$= \mathbf{W}(z,k)\mathbf{H}(z)\mathbf{s}(k) = \mathbf{C}(z,k)\mathbf{s}(k) \tag{3}$$

where $\mathbf{y}(k) \in \mathbb{C}^N$ is the output vector of estimated source signals with entries $y_i(k) = \sum_{j=1}^{N} W_{ij}(z,k)x_j(k)$, $\mathbf{W}(z,k)$ is the unmixing temporal network with entries $W_{ij}(z,k) = \sum_{p=0}^{L} w_{ijp}(k)z^{-p}$. The task is to adjust $\mathbf{W}(z,k)$ such that $\lim_{k\to\infty} \mathbf{C}(z,k) = \mathbf{P\Lambda D}(z)$, where $\mathbf{P} \in \mathbb{R}^{N\times N}$ is a permutation matrix, $\mathbf{\Lambda} \in \mathbb{R}^{N\times N}$ is a non-singular diagonal scaling matrix, and $\mathbf{D}(z)$ is a diagonal matrix whose iith entry is $\sum_p c_{ip}z^{-p}$, c_{ip} is a complex scalar weighting, and p is an integer delay value. This formulation makes the i.i.d. assumption in the deconvolution context unnecessary for convolutive BSS. Taking advantage of the scale and permutation indeterminacy of the conventional BSS methods [9], we assume that the cyclic correlation matrix of the sources at lag $\tau = 0$ follows $\mathbf{R}_{\mathbf{s}}^{\beta_i}(k)[0] = E\{e^{J\beta_i k}\mathbf{s}(k)\mathbf{s}^{\mathrm{H}}(k)\} = \mathbf{I}'_i$, where $(\cdot)^{\mathrm{H}}$ denotes the Hermitian transpose operator, $\mathbf{I}'_i$ has entries as follows:

$$[\mathbf{I}'_i]_{l,g} = \begin{cases} 1 & \text{if } l \in \{1,2,\ldots,N\},\ g = l = i \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

3. THE PROPOSED ALGORITHM

3.1. Separation of instantaneous mixtures of cyclostationary sources

The wide sense cyclostationarity assumption implies that the cyclic correlation function of the sources satisfies [7]:

$$\langle e^{J\beta_i k} s_i(k+\tau)s_j^*(k)\rangle = 0 \quad \text{if } i \neq j \tag{5}$$

$$\langle e^{J\beta_j k} s_i(k+\tau)s_i^*(k)\rangle = 0 \quad \text{if } \beta_i \neq \beta_j \tag{6}$$

$$\langle e^{J\beta_i k} s_i(k)s_i^*(k)\rangle > 0 \quad \forall i \tag{7}$$

where $J = \sqrt{-1}$, the superscript * denotes complex conjugation, $\langle\cdot\rangle$ denotes the time averaging operator, and β_i is a non-zero cycle frequency of source i. It is worth noting that we use $\langle\cdot\rangle$ and $E(\cdot)$ to denote time-averages for discrete- and continuous-time operations, respectively, as the context demands [10]. The equivalence has been shown in Reference [8]. Invoking these properties into the Kullback–Leibler divergence criterion, our cost function is defined as (see References [9, 11])

$$\rho(\mathbf{W}(k)) = -\log\det(\mathbf{W}(k)) - \sum_{i=1}^{N} \log p_i(y_i(k)) + \frac{1}{2}\{\mathrm{Tr}(\tilde{\mathbf{R}}_{\mathbf{y}}^{\boldsymbol{\beta}}(k)) - \log\det(\tilde{\mathbf{R}}_{\mathbf{y}}^{\boldsymbol{\beta}}(k)) - N\} \tag{8}$$

where $\mathrm{Tr}(\cdot)$ and $\det(\cdot)$ are, respectively, the trace and determinant operators, and $p_i(y_i(k))$ is an appropriately chosen probability density function (pdf) of $\mathbf{y}(k)$ which is derived from the past sequence of $\mathbf{x}(k)$ by the separation matrix $\mathbf{W}(k)$. The term $\tilde{\mathbf{R}}_{\mathbf{y}}^{\boldsymbol{\beta}}(k)$ is defined as

$$\tilde{\mathbf{R}}_{\mathbf{y}}^{\boldsymbol{\beta}}(k) = \sum_{i=1}^{N} \mathbf{R}_{\mathbf{y}}^{\beta_i}(k) \tag{9}$$

where $\mathbf{R}_{\mathbf{y}}^{\beta_i}(k) = \langle e^{J\beta_i k}\mathbf{y}(k)\mathbf{y}^{T}(k)\rangle$ is the output cyclic correlation matrix for the ith cycle frequency which is required to satisfy $\lim_{k\to\infty} \mathbf{R}_{\mathbf{y}}^{\beta_i}(k) = \mathbf{I}_i'$, where the elements of $\mathbf{I}_i'$ take the form in (4) and time lag in $\mathbf{R}_{\mathbf{y}}^{\beta_i}(k)$ is omitted for simplicity. At convergence,

$$\lim_{k\to\infty} \tilde{\mathbf{R}}_{\mathbf{y}}^{\boldsymbol{\beta}}(k) = \mathbf{I} \tag{10}$$

Iterative learning can be used to minimize the cost function in (8). Applying the natural gradient rule [12] together with a cyclic decorrelation operation, we obtain a new learning rule as (see the appendix)

$$\begin{aligned}\mathbf{W}(k+1) = \mathbf{W}(k) + \mu(k)[\mathbf{I} - \mathbf{f}(\mathbf{y}(k))\mathbf{y}^{T}(k) \\ + \mathbf{I} - \tilde{\mathbf{R}}_{\mathbf{y}}^{\boldsymbol{\beta}}(k)]\mathbf{W}(k)\end{aligned} \tag{11}$$

where $\mu(k)$ is the learning rate and $\mathbf{f}(\cdot)$ is a vector of non-linear activation functions. For simplicity, we call the algorithm in (11) the CSNGA algorithm corresponding to the natural gradient algorithm (NGA). Observing the learning rule in (11), we see that the proposed algorithm has a similar form to the equivariant adaptive algorithm proposed by Cardoso and Laheld [13] and equivariant non-stationary algorithm proposed by Choi *et al*. [14]. More specifically, comparing the term $\mathbf{I} - \tilde{\mathbf{R}}_{\mathbf{y}}^{\boldsymbol{\beta}}(k)$ in (11) and its counterpart in References [13, 14], it is clear that rather than incorporating the cyclostationary statistics as in (11), a whitening process with the form of $\mathbf{I} - \mathbf{y}\mathbf{y}^{T}$ was directly incorporated in the learning process in Reference [13], and a form of $\mathbf{I} - \Pi^{-1}\mathbf{y}\mathbf{y}^{T}$ was employed for the non-stationary statistics in Reference [14], where the elements of the diagonal matrix Π can be estimated by the variance of $\mathbf{y}$.

3.2. Separation of convolutive mixtures of cyclostationary sources

For convolutive BSS, from the algebraic point of view, they have equivalent mathematical models to the instantaneous cases, only being different in the description (IIR filter, FIR filter, $\mathscr{Z}$-transform, wavelets, other transforms) of the physical phenomena [12]. This means, for convolutive mixtures, the complicated problem can be simplified by applying such transforms. In terms of this fact, it is straightforward to use the $\mathscr{Z}$-transform and define the following cost function for convolutive mixtures of cyclostationary sources (see Reference [15]):

$$\rho(\mathbf{W}(z,k)) = \rho_\gamma(\mathbf{W}(z,k)) + \rho_\beta(\mathbf{W}(z,k)) \tag{12}$$

where $\rho_\gamma(\mathbf{W}(z,k))$ and $\rho_\beta(\mathbf{W}(z,k))$ are, respectively,

$$\rho_\gamma(\mathbf{W}(z,k)) = -\sum_{i=1}^{N} \log p_i(y_i(k)) - \frac{1}{2\pi J}\oint \log|\det\mathbf{W}(z,k)|z^{-1}\,dz$$

$$\rho_\beta(\mathbf{W}(z,k)) = \tfrac{1}{2}\{\mathrm{Tr}(\tilde{\mathbf{R}}_{\mathbf{y}}^{\boldsymbol{\beta}}(k)) - \log\det(\tilde{\mathbf{R}}_{\mathbf{y}}^{\boldsymbol{\beta}}(k)) - N\}$$

Here, we assume both $\mathbf{H}(z)$ and $\mathbf{W}(z,k)$ are stable with no zero eigenvalues on the unit circle $|z| = 1$. To obtain a learning algorithm which minimizes the cost function (12), we follow the derivation of the algorithm presented by Amari *et al*. [12] and calculate the infinitesimal increment $d\rho(\mathbf{W}(z,k))$ corresponding to an increment $d\mathbf{W}(z,k)$. Introducing the Lie group operation and Riemanian structure on the manifold of FIR filters $\mathscr{M}(L)$, $d\mathbf{X}(z,k) = d\mathbf{W}(z,k) \circledast \mathbf{W}^{\dagger}(z)$, where $\circledast$ and † are the multiplication and inverse operation defined in the manifold $\mathscr{M}(L)$, $\mathrm{d}\rho_\gamma(\mathbf{W}(z,k))$ takes the following form by simple algebraic and differential calculus [12]:

$$\mathrm{d}\rho_\gamma(\mathbf{W}(z,k)) = \mathbf{f}^{\mathrm{T}}(\mathbf{y}(k))\,\mathrm{d}\mathbf{X}(z,k)[\mathbf{y}(k)] - \mathrm{Tr}(\mathrm{d}\mathbf{X}(z,k)) \quad (13)$$

Notice that $\mathbf{y}(k) = \mathbf{W}(z,k)\mathrm{x}(k)$, $\tilde{\mathbf{R}}_{\mathbf{x}} = \langle \mathbf{x}(k)\mathbf{x}^{\mathrm{T}}(k)\rangle$, and the property $\mathrm{d}\mathbf{y}(k) = \mathrm{d}\mathbf{X}(z,k)\mathbf{y}(k)$, we have (see appendix)

$$\mathrm{d}\{\log\det(\tilde{\mathbf{R}}_{\mathbf{y}}^{\boldsymbol{\beta}}(k))\} = \mathrm{d}\left\{N\log\left(\sum_{i=1}^{N} \mathrm{e}^{\mathrm{J}\beta_i k}\right)\right\} + \mathrm{d}\{\log\det\tilde{\mathbf{R}}_{\mathbf{x}}\} + 2\,\mathrm{Tr}\{\mathrm{d}\mathbf{W}(z,k)\mathbf{W}^{-1}(z,k)\} \quad (14)$$

$$\mathrm{d}\{\mathrm{Tr}(\tilde{\mathbf{R}}_{\mathbf{y}}^{\boldsymbol{\beta}}(k))\} = 2\left(\sum_{i=1}^{N} \mathrm{e}^{\mathrm{J}\beta_i k}\right)\langle \mathbf{y}^{\mathrm{H}}(k)\,\mathrm{d}\mathbf{X}(z,k)\mathbf{y}(k)\rangle \quad (15)$$

where $\mathrm{d}\mathbf{X}(z,k)$ is denoted as $\sum_{p=0}^{L} \mathrm{d}\mathbf{X}_p(k)z^{-p}$. Notice that $\mathrm{d}\left\{N\log\left(\sum_{i=1}^{N} \mathrm{e}^{\mathrm{J}\beta_i k}\right)\right\}$ and d{log det $\tilde{\mathbf{R}}_{\mathbf{x}}$} do not depend on the weight matrix $\mathbf{W}(z,k)$, $\mathrm{d}\left\{\log\det\left(\tilde{\mathbf{R}}_{\mathbf{y}}^{\boldsymbol{\beta}}(k)\right)\right\}$ can be reduced to 2 tr $\mathrm{d}\mathbf{W}(z,k)\mathbf{W}^{-1}(z,k)\}$, therefore we have

$$\begin{aligned}\mathrm{d}\rho_\beta(\mathbf{W}(z,k)) = &-\mathrm{Tr}\{\mathrm{d}\mathbf{W}(z,k)\mathbf{W}^{-1}(z,k)\}\\ &+\left(\sum_{i=1}^{N} \mathrm{e}^{\mathrm{J}\beta_i k}\right)\langle \mathbf{y}^{\mathrm{T}}(k)\,\mathrm{d}\mathbf{X}(z,k)\mathbf{y}(k)\rangle\end{aligned} \quad (16)$$

From (13) and (16), the partial derivation of $\rho(\mathbf{W}(z,k))$ with respect to $\mathbf{X}(z,k)$ leads to

$$\begin{aligned}\frac{\partial\rho(\mathbf{W}(z,k))}{\partial\mathbf{X}_p(k)} = &\{\mathbf{f}(\mathbf{y}(k))\mathbf{y}^{\mathrm{T}}(k-p) - \mathbf{I}\delta_p\}\\ &+\left\{-\mathbf{I}\delta_p + \sum_{i=1}^{N}\langle \mathrm{e}^{\mathrm{J}\beta_i k}\mathbf{y}(k)\mathbf{y}^{\mathrm{T}}(k-p)\rangle\right\}\end{aligned} \quad (17)$$

The truncated form of $\mathrm{d}\mathbf{X}(z,k)$ on the FIR filter manifold $\mathscr{M}(L)$, i.e. $[\mathrm{d}\mathbf{W}(z,k)\mathbf{W}^{-1}(z,k)]_L$, implies that the partial derivative of $\rho(\mathbf{W}(z,k))$ with respect to $\mathbf{W}(z,k)$ can be obtained according to the partial derivative with respect to $\mathbf{X}(z,k)$, this leads to a new learning algorithm for the separation of convolutive mixtures of cyclostationary signals

$$\begin{aligned}\mathbf{W}_p(k+1) = \mathbf{W}_p(k) + \mu(k)\{&\mathbf{W}_p(k) - \hat{\mathbf{Q}}_q(k)\\ &+\mathbf{W}_p(k) - \mathbf{f}(\mathbf{y}(k))\mathbf{u}_\gamma^{\mathrm{T}}(k)\}\end{aligned} \quad (18)$$

where $\mathbf{u}_\gamma(k) = \sum_{q=0}^{L} \mathbf{W}_{L-q}^{\mathrm{T}}(k)\mathbf{y}(k-p-q)$, $\hat{\mathbf{Q}}_q(k) = \sum_{q=0}^{L} \mathbf{W}_{L-q}^{\mathrm{T}}(k)\hat{\mathbf{R}}_{\mathbf{y}}^{\boldsymbol{\beta}}(k)[q]$, and $\hat{\mathbf{R}}_{\mathbf{y}}^{\boldsymbol{\beta}}(k)[q]$, equivalent to the last term in (17), can be estimated by a sample filtering approach, $\hat{\mathbf{R}}_{\mathbf{y}}^{\boldsymbol{\beta}}(k)[p] = (1-\eta_0)\hat{\mathbf{R}}_{\mathbf{y}}^{\boldsymbol{\beta}}(k)[p] + \eta_0\sum_{i=1}^{N} \mathrm{e}^{\mathrm{J}\beta_i k}\mathbf{y}(k)\mathbf{y}^{\mathrm{T}}(k-p)$, where η_0 is a forgetting factor. For simplicity, we call the algorithm in (18) the ConvCSNGA algorithm corresponding to the natural gradient algorithm for convolutive mixtures (ConvNGA).

The equilibrium points of the proposed learning algorithm satisfy

$$\lim_{k\to\infty}\left\langle \mathbf{I} - \sum_{i=1}^{N} \mathrm{e}^{\mathrm{J}\beta_i k}\mathbf{y}(k)\mathbf{y}^{\mathrm{T}}(k)\right\rangle = \mathbf{0} \quad (19)$$

$$\lim_{k\to\infty}\left\langle \sum_{i=1}^{N} e^{J\beta_i k}\mathbf{y}(k)\mathbf{y}^{T}(k-p)\right\rangle = \mathbf{0} \tag{20}$$

$$\lim_{k\to\infty}\langle \mathbf{f}(y(k))\mathbf{y}^{T}(k-p)\rangle = \mathbf{0} \tag{21}$$

$$\lim_{k\to\infty}\langle \mathbf{I}-\mathbf{f}(\mathbf{y}(k))\mathbf{y}^{T}(k)\rangle = \mathbf{0} \tag{22}$$

where $p = 1, 2, \ldots, L$. This implies that the proposed learning algorithm has a drawback that it forces the output signals to have nearly flat frequency spectra, which however can be mitigated by following the non-holonomic constraint, that is, using $\mathbf{W}_p(k)\boldsymbol{\Lambda}(k)$ instead of $\mathbf{W}_p(k)$ in (18), where $\boldsymbol{\Lambda}(k) = \mathrm{diag}\{\mathbf{f}(\mathbf{y}(k))\mathbf{y}^{T}(k)\}$ (see Reference [12] for more details).

In implementation, the infinite system is approximated by the FIR filter given $\mathbf{y}(k) = \sum_{p=0}^{L} \mathbf{W}_p(k)\mathbf{x}(k-p)$ and a similar approximation to the updating equation for the general case that all signals and coefficients are complex-valued yields

$$\begin{aligned}\mathbf{W}_p(k+1) = \mathbf{W}_p(k) + \mu(k)\{&\mathbf{W}_p(k) - \hat{\mathbf{Q}}_{p+q}(k-L)\\ &+ \mathbf{W}_p(k) - \mathbf{f}(\mathbf{y}(k-L))\mathbf{u}_{\gamma}^{H}(k-p)\}\end{aligned} \tag{23}$$

It is necessary to use split-complex non-linearities in the complex case, e.g. $f_i(y_i(k)) = \tanh(y_{i\mathrm{R}}(k)) + J\tanh(y_{i\mathrm{I}}(k))$, where $y_{i\mathrm{R}}(k)$ and $y_{i\mathrm{I}}(k)$ are, respectively, the real and imaginary parts of $y_i(k)$, for the sources with super-Gaussian distributions, or $f_i(y_i(k)) = y_i(k)|y_i(k)|^2$ for the sub-Gaussian case [9, 13]. In practice, the output cyclic operation term is estimated on-line using an exponentially weighted moving average of the statistics with the filtering output data. The algorithm requires approximately $4MN(N+1)(L+1) + 2N$ multiplications per time instant, and it requires approximately $((M+1)(N+1)N + M)(L+1)$ memory locations to implement. A normalized step size is used in practice $\mu(k) = \mu/\{\rho + \sum_{p=0}^{L} |\mathbf{y}(k-p)|^2\}$, where ρ is a positive value which avoids an explosive growth of step size. Another implementation problem is the selection of cyclic frequency. The cyclic frequency detection will be discussed in Section 4.

3.3. Stability and convergence analysis of the learning algorithm

To analyse the stability of the learning algorithm (18), we follow the method in References [16, 17] and alternatively consider the continuous-time learning algorithm for updating $\mathbf{X}_p$:

$$\frac{d\mathbf{X}_p}{dt} = \mu(\delta_p\mathbf{I} - \langle \mathbf{f}(\mathbf{y}(t))\mathbf{y}^{T}(t-p)\rangle + \delta_p\mathbf{I} - \sum_{i=1}^{N}\langle e^{J\beta_i k}\mathbf{y}(t)\mathbf{y}^{T}(t-p)\rangle) \tag{24}$$

where $p = 0, 1, \ldots, L$. By considering expectation in Equation (24), we can examine the behaviour of the averaged system

$$\frac{d\mathbf{X}_p}{dt} = \mu(\delta_p\mathbf{I} - E\{\mathbf{f}(\mathbf{y}(t))\mathbf{y}^{T}(t-p)\} + \delta_p\mathbf{I} - \sum_{i=1}^{N} E\{e^{J\beta_i k}\mathbf{y}(t)\mathbf{y}^{T}(t-p)\} \tag{25}$$

Taking a variation $\delta\mathbf{X}_p$ of $\mathbf{X}_p$, we have

$$\frac{d\delta\mathbf{X}_p}{dt} = -\mu(E\{\mathbf{f}'_{\boldsymbol{\beta}}(\mathbf{y})\delta\mathbf{y}\mathbf{y}^{T}(t-p) + \mathbf{f}_{\boldsymbol{\beta}}(\mathbf{y})\delta\mathbf{y}^{T}(t-p)\} \tag{26}$$

where $\mathbf{f}_{\beta}(\mathbf{y}) = \mathbf{f}(\mathbf{y}(t)) + e^{J\beta k}\mathbf{y}(t)$ and $\mathbf{f}'_{\beta}(\mathbf{y})$ is the derivation of $\mathbf{f}_{\beta}(\mathbf{y})$ with respect to $\mathbf{y}$, $e^{J\beta k} \triangleq \sum_{i=1}^{N} e^{J\beta_i k}$, $\delta\mathbf{y}(k-p) = [\delta\mathbf{X}(z)]\mathbf{y}(k-p) = \sum_{j=0}^{\infty} \delta\mathbf{X}_p\mathbf{y}(k-p-j)$. Due to the spatially independent assumption of the source signals and the conditions in (19)–(22), if we further suppose the sources are temporally identical and independent, then (26) can be reduced to the following equation by omitting the last term:

$$\frac{\mathrm{d}\delta\mathbf{X}_p}{\mathrm{d}t} = \begin{cases} -\mu(E\{\mathbf{f}'_{\beta}(\mathbf{y})\delta\mathbf{X}_p\mathbf{y}(t-p)\mathbf{y}^{\mathrm{T}}(t-p)\}), & p \neq 0 \\ -\mu(E\{\mathbf{f}'_{\beta}(\mathbf{y})\delta\mathbf{X}_0\mathbf{y}\mathbf{y}^{\mathrm{T}}\} + \delta\mathbf{X}_0), & p = 0 \end{cases} \tag{27}$$

Let $\tilde{m}_i = E\{f'_{\beta}(y_i)y_i^2\}$, $\tilde{\kappa}_i = E\{f'_{\beta,i}(y_i)\}$, and $\tilde{\sigma}_i^2 = E\{|y_i|^2\}$. The stability conditions for (27) can be formulated as

$$\tilde{m}_i + 1 > 0 \tag{28}$$

$$\tilde{\kappa}_i > 0 \tag{29}$$

$$\tilde{\sigma}_i^2\tilde{\sigma}_j^2\tilde{\kappa}_i\tilde{\kappa}_j > 1 \tag{30}$$

The conditions are similar to the ones by Zhang *et al.* [16] for blind source separation of the convolutive case and by Amari *et al.* [17] for the instantaneous case, which can be satisfied by choosing a suitable non-linear activation function plus a linear function with cycle frequency coefficients.

4. CYCLE FREQUENCY DETERMINATION FOR BSS

The CSNGA algorithm in (11) and the ConvCSNGA algorithm in (18), both require knowledge of the cycle frequency of each signal. In practice, however, it is very unlikely that this information is available *a priori*, especially since the set of cycle frequencies grows as the number of source signals increases. To overcome this drawback, it is assumed that at least one cycle frequency β_i is known, and the algorithm is modified so that β_i is used to construct the output cyclic correlation matrix $\hat{\mathbf{R}}_{\mathbf{y}}^{\beta}(k)$ and $\hat{\mathbf{R}}_{\mathbf{y}}^{\beta}(k)[p]$, while Equations (10), (19) and (20) are still satisfied. Effectively, this amounts to replacing the sum of all correlation matrices obtained for every individual cycle frequency with the sum of an equivalent number of matrices generated by permutation operations to the matrices, which leads to similar rates of convergence and steady-state characteristics and ensures that all estimated sources are decorrelated in the cyclostationary sense, provided that at least one cycle frequency is known *a priori* [18]. In practice, this last condition is a reasonable hypothesis, since the cycle frequency of a signal can be estimated with techniques such as those outlined in Reference [19], where the cycle frequency of the ith source signal is estimated using the heart instantaneous frequency (HIF) approach proposed by Barros and Ohnishi [19]. This method was developed for the extraction of the instantaneous heart rate from electrocardiogram (ECG) signals, which are non-stationary and whose frequency varies over time due, for instance, to changes in respiration. The HIF method is a block estimation technique, whose implementation involves selecting a suitable signal $z(k)$ from which the instantaneous frequency is determined. Although the obvious choice for $z(k)$ would be any of the observed signals, the presence of all cycle frequencies in the ith mixture $x_i(k)$ implies that the algorithm may extract different frequencies at different times, a problem that becomes more conspicuous when the mixing channel is time-varying. Alternatively, we propose a sequential method to estimate the cycle frequency from one of the extracted sources using the adaptive scheme in Figure 1. By modifying the HIF approach in Reference [19], a sequential algorithm for cycle frequency determination is concluded as follows [18]:

(i) Select one output signal and acquire a block of length B as the data become available.
(ii) Calculate the spectrogram $S(k, f)$ of $z(k)$ in terms of the acquired block samples.
(iii) Find the frequency value at which $S(k, f)$ attains its maximum in a given frequency interval $[\delta(k^-) - \gamma, \delta(k^-) + \gamma]$, where $\delta(k^-)$ is the value of the driver at the previous time, and the frequency range searched is limited by γ.
(iv) Divide the original signal $z(k)$ into blocks, to give the signals $z_\Omega(k)$, where Ω is a short time interval.
(v) Calculate the filtered version $\tilde{z}_\Omega(k)$ of $z_\Omega(k)$ using a band-pass filter $u(k)$ with centre frequency $\overline{\delta(k)}$.
(vi) Evaluate the angular instantaneous frequency for each $\tilde{z}_\Omega(k)$ from $\omega(k) = \phi(k+1) - \phi(k)$, $\phi(k) = \tan^{-1}(-H[\tilde{z}_\Omega(k)]/\tilde{z}_\Omega(k))$, where $H[\tilde{z}_\Omega(k)]$ is the Hilbert transform of the signal $\tilde{z}_\Omega(k)$.
(vii) Evaluate the frequency from $\omega(k)$, using $f(k) = \omega(k)/2\pi$. Then, the estimated cyclic frequency is given by $\hat{\beta}_{\mathrm{HIF}}(k) = 2f(k)$.
(viii) Use the resulting cycle frequency over this time interval to separate the sources with the ConvCSNGA algorithm.
(ix) Repeat (i)–(vii) for every new block of data.

Steps (ii)–(vii) are actually specified in Figure 1 termed as 'cycle frequency determination'. In our implementation, the windowing function $h(k)$ is chosen to be the Hanning window of length 256, the band-pass filter $u(k)$ is a second-order digital filter based upon a Butterworth prototype, and the frequency range searched is limited by $\gamma = 17$ frequency samples. Finally, the data block size B is set to 100 samples. The method presented above implies implicitly that the source signals have distinct cyclic frequencies and that the noise is at a low level, more details are found in Reference [18] for the practical problems such as identical cycle frequencies and the robustness of the algorithm to additive noise.

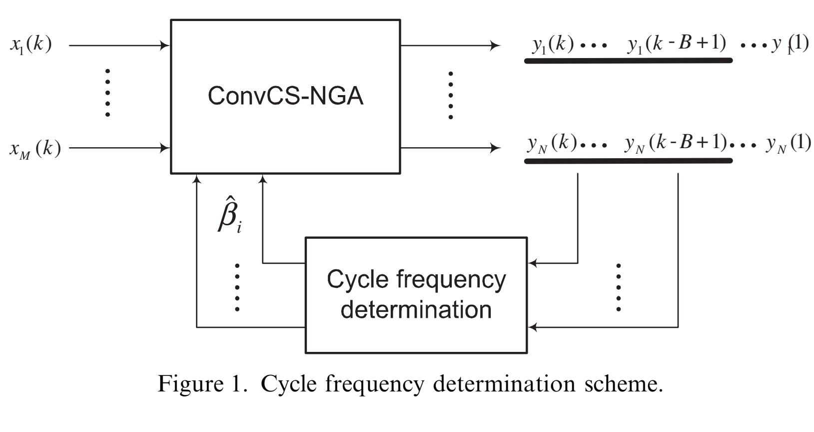

Figure 1. Cycle frequency determination scheme.

5. SIMULATION STUDY

In this section, we examine the performance of the proposed algorithm by simulations both for instantaneous mixtures and convolutive mixtures. For convolutive mixtures, we will deal with two kinds of data, namely, artificial convolutive mixtures and the data of real recordings.

5.1. Instantaneous mixtures

We use the ICALAB Toolbox [20] to compare the performance of the proposed algorithm for a four-input-four-output system ($N = M = 4$) with several other existing BSS algorithms which are, respectively, joint approximate diagonalization of eigen matrices with optimized numerical procedures (JADEop) and with time delays (JADETD), online natural gradient algorithm (NG-OL), fix-point ICA algorithm (FPICA), second-order blind identification algorithm (SOBI), algorithm for multiple unknown source extraction based on EVD (AMUSE), and self-adaptive natural gradient algorithm with non-holonomic constraints (SANG) (see References [12, 20] and the reference therein [21–30]). It should be noted that JADEop is a modified version of the initial work by Cardoso [26] (see Reference [20]). The test scheme is designed as in Figure 2. The four source benchmarks which all display same form of cyclostationarity are, respectively, sinusoid signals, BPSK modelled signals, first-order autoregressive (AR) Gaussian process modelled by sinusoid signals, and speech signals. More specifically, the AR signals are obtained from $s_i(k) = n_i(k)/(1 - a_i z^{-1})$ $(i = 1, \ldots, 4)$ modulated by sinusoids $\cos \alpha_i$ with the values of α_i listed in Table I. The four coefficients a_i of the first-order AR signals are all $0.9 \cos(0.5)$, and $n_i(k)$ are all the Gaussian process with standard variance 0.1. The speech signals are from the recordings of reading four words 'some', 'pick', 'sing' and 'date' which are all sampled at 10 kHz and 16-bits with a duration of 0.3 s. The cycle frequencies of the four benchmarks are listed in Table I. It should be noted that the cycle frequencies of speech signals were approximately calculated

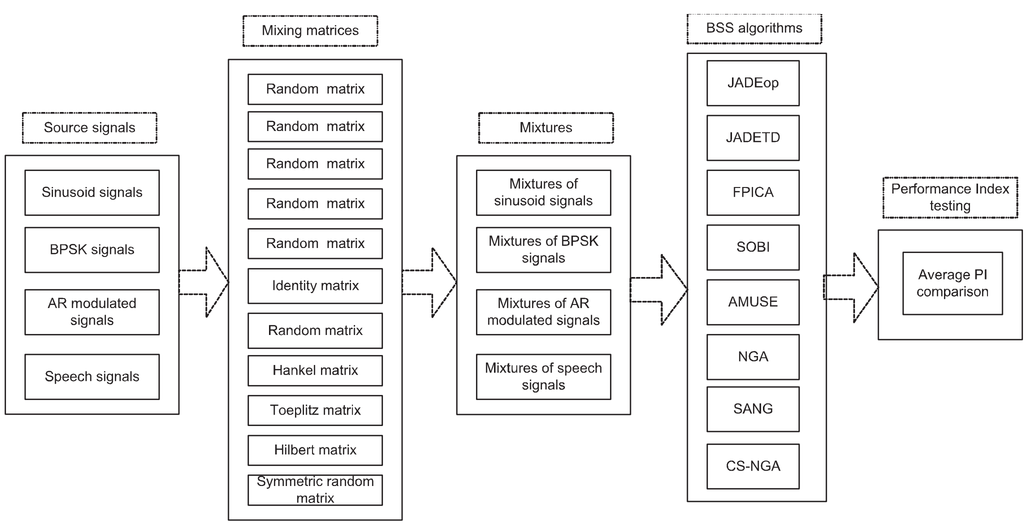

Figure 2. PI testing model of the proposed algorithm (CSNGA) as compared with other BSS algorithms (JADEop, JADETD, FPICA, SOBI, AMUSE, NGA, SANG) for instantaneous mixtures.

Table I. Cycle frequencies of the four benchmarks.

Benchmarks	Cycle frequencies of the four source signals			
Sinusoid signals	$(20\pi)^{-1}$	$(10\pi)^{-1}$	$3(10\pi)^{-1}$	$9(20\pi)^{-1}$
BPSK modelled signals	$(10\pi)^{-1}$	$(5\pi)^{-1}$	$(4\pi)^{-1}$	$2(5\pi)^{-1}$
AR modelled signals	$(10\pi)^{-1}$	$(4\pi)^{-1}$	$3(10\pi)^{-1}$	$9(20\pi)^{-1}$
Speech signals (approximate)	180	200	150	170

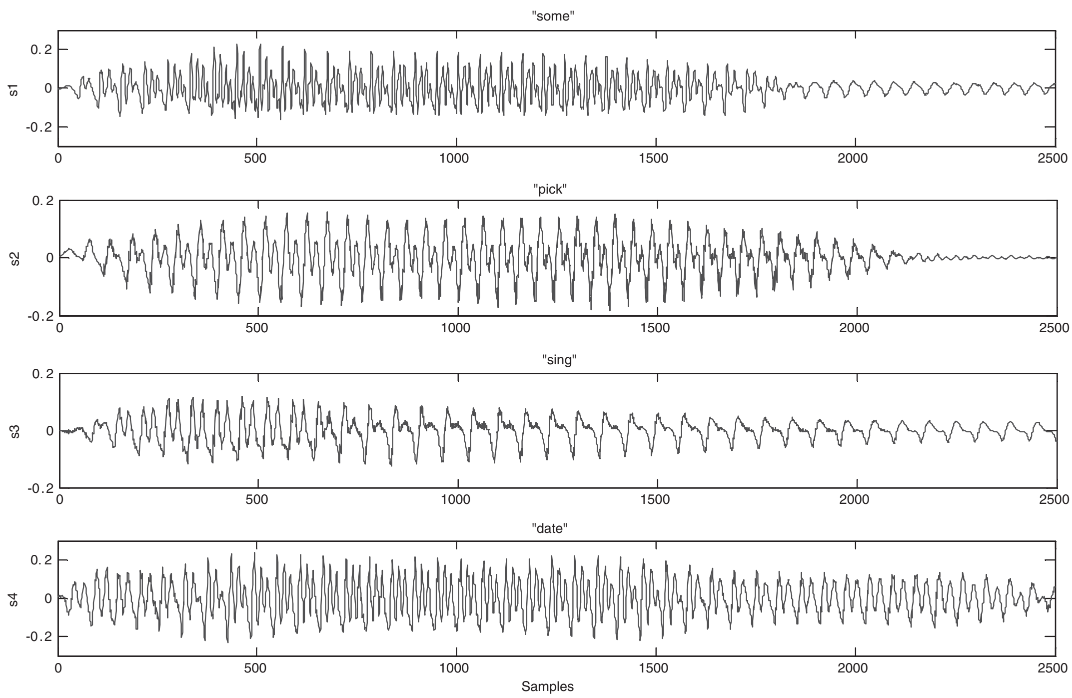

Figure 3. Speech signals with cyclic properties.

according to the individual time scale and periodicity of the voiced part of each speech signal (Figure 3). The determinants and the condition values of the 11 mixing matrices are listed in Table II. The normalized performance index (PI) which is often used to evaluate the performance of BSS algorithms is defined as

$$\mathrm{PI}(k) = \frac{1}{m}\sum_{i=1}^{m}\left\{\sum_{j=1}^{m}\frac{|p_{ij}|^2}{\max_q|p_{iq}|^2} - 1\right\} + \frac{1}{m}\sum_{j=1}^{m}\left\{\sum_{i=1}^{m}\frac{|p_{ij}|^2}{\max_q|p_{qj}|^2} - 1\right\} \quad (31)$$

where p_{ij} are the entries of the global system $\mathbf{P} = \mathbf{WH}$. Generally speaking, a lower value indicates a better performance. In the test, we use the default parameters which are already tuned approximate optimum values for typical data, and the performance of the algorithms is evaluated by the average PIs of five trials for each algorithm which is shown in Figure 4. From

Table II. Determinants and condition values of the mixing matrices.

Mixing matrices	H_1	H_2	H_3	H_4	H_5	H_6
Determinants	−0.46	0.26	0.55	−1.74	0.24	1
Condition values	10.74	9.94	4.93	2.28	5.63	1
Mixing matrices	H_7	H_8	H_9	H_{10}	H_{11}	
Determinants	0.75	1.0e−3	0.41	1.33e−7	3.33e−4	
Condition values	5.04	94.75	5.32	1.98e+4	10.08	

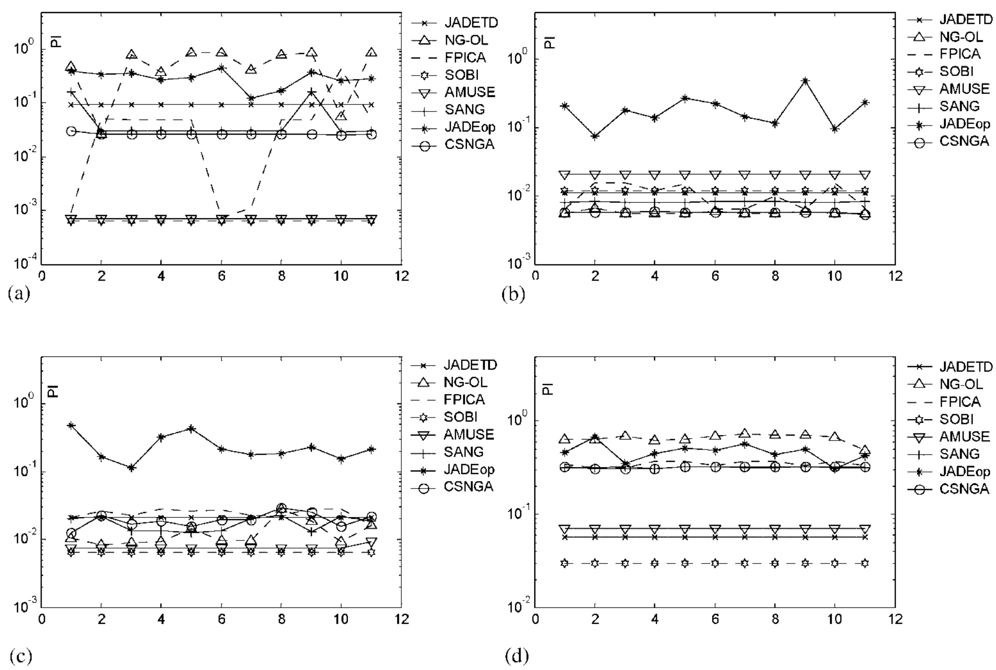

Figure 4. PI comparison of the proposed algorithm (CSNGA) with several other BSS algorithms (JADETD, NG-OL, FPICA, SOBI, AMUSE, SANG, JADEop) for the separation of four kinds of signals with cyclostationary properties. The four signals are, respectively, sinusoid signals, BPSK modelled signals, AR modelled signals and speech signals.

Figure 4, we see that although the proposed algorithm is based on high-order statistics which is the same as in the natural gradient algorithm and can be regarded as a new member of the natural gradient algorithms, incorporation of the second-order cyclostationary statistics shows a superior performance to that of NG-OL and SANG algorithms for the separation of the first three source signals (see Figure 4 (a)–(c)). The proposed algorithm has a similar performance for the separation of the sine wave sources as the second-order methods such as AMUSE and SOBI, however, the high-order statistics ICA algorithm (such as JADEop and NG-OL) fail to reconstruct such sources because they are dependent (see Figure 4(a)). It should be noted that other high-order statistics based methods such as JADETD and FPICA, similarly, do not perform well for such sources either. However, there exists slight difference (even inconsistent results) in performance measurement due to the various learning mechanism in these methods (see Reference [20] for more details). From Figure 4(d), it can be seen that considering the cyclic properties of the sources also generates a better performance as compared to the NG-OL for the separation of typical speech signals with cycle-like properties, however, the performance is not as good as that of the SANG algorithm. This is not unexpected since the speech signals are non-stationary and their average magnitudes change rapidly and the SANG algorithm introduces non-holonomic constraints in the NG learning process so that it can adapt to rapid changes in the magnitudes of the source signals. In our simulations, the proposed algorithm without such constraints can however generate a similar separation performance as the SANG algorithm.

5.2. Convolutive mixtures

5.2.1. Artificially convolutive mixtures. In the case of artificially convolved mixtures, a two-input-two-output (TITO) system ($N = M = 2$) is considered for simplicity. The resemblance between the original and the reconstructed source waveforms is measured by their mean squared difference

$$\varepsilon^2(\text{dB}) = 10\log\left\{(1/N)\sum_{i=1}^{N} E[|y_i(k) - s_i(k)|^2]\right\} \tag{32}$$

(assuming the signals are zero-mean and unit-variance). First, we use sinusoid signals to test the performance of the proposed algorithm, where the sinusoid signal can be treated as a degenerate case of cyclostationary signals, we appreciate that such a signal will only excite one of the frequency components of a general convolutive system. The two source signals are set, respectively, to be $s_1(i) = \sin(0.4i)$, $s_2(i) = \sin(0.9i)$. The two sources are convolved by FIR filters with entries $\mathrm{H}(1,:,1) = [1\ 1\ -0.75\ 0.9]$, $\mathrm{H}(2,:,1) = [0.5\ -0.3\ 0.2\ 0.2]$, $\mathrm{H}(1,:,2) = [-0.2\ 0.4\ 0.7\ 0.2]$, $\mathrm{H}(2,:,2) = [0.2\ 1\ 0\ 0.9]$. The other parameters are $\mu = 0.004$, and $\rho = 0.1$. The two cyclic frequencies are, respectively $(5\pi)^{-1}$ and $9(20\pi)^{-1}$. It should be noted that, in general, changing filter length L would change the performance of the separation quality [5]. It has been shown that if we increase L, we may obtain an improved separation performance as compared with a shorter length, however, the performance will also suffer from the length of the data samples being used when increasing L, and it may degraded seriously when the data length is limited. Additionally, for the time-domain implementation, increasing L indicates a significant increase of the filter coefficients to be estimated. This will lead to a problem of much more computational complexity as compared to that of the frequency domain implementation. In the following simulations, we use $L = 32$ if it is not specified. The separation results are shown in Figure 5, and the performance indices are, respectively, -8.36 and -4.35 dB. In the second simulation (see Figure 6), two BPSK signals modulated by sinusoids of carrier frequencies $(5\pi)^{-1}$ and $9(20\pi)^{-1}$, which can be treated as the other general sets of signals with cyclic property, are chosen as the source signals. The mixing matrix and the parameters are identical to those in the first simulation. The performance indices are, respectively, -14.95 and -4.45 dB. In both simulations, the cyclic frequencies are chosen to be the same as the carrier frequencies and the step size is unchanged for two simulations (allowed for comparison). Both simulation results indicate that the proposed algorithm has improved performance over the natural gradient

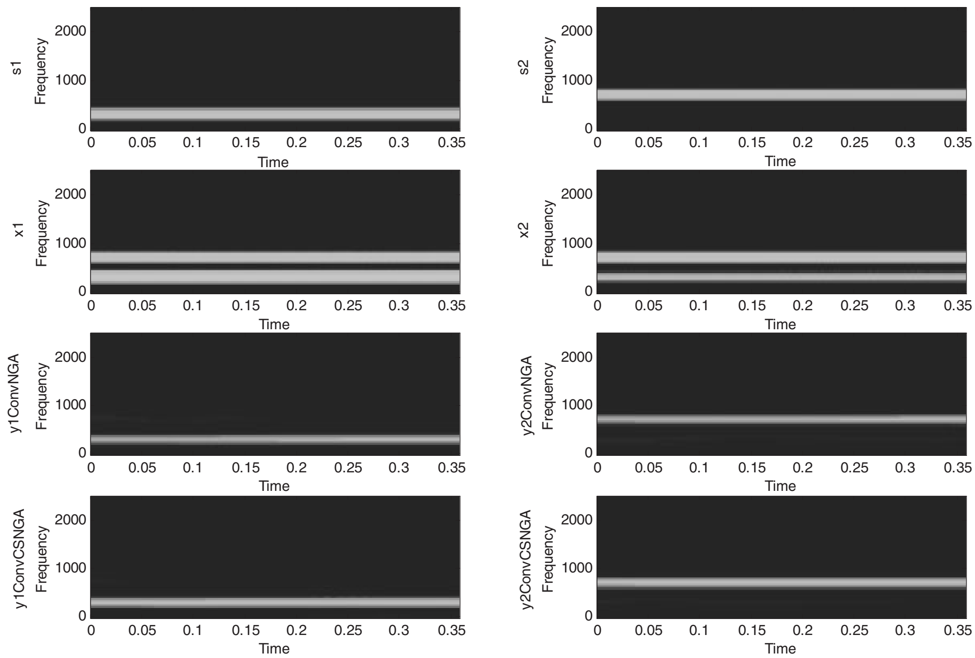

Figure 5. Separation of convolutive mixtures of two sinusoid signals: $s1$ and $s2$ are two sources, $x1$ and $x2$ are the two convolutive mixtures, y1ConvNGA and y2ConvNGA are the recovered signals by the natural gradient algorithm, y1ConvCSNGA and y2ConvCSNGA are the recovered signals by the proposed method.

algorithm in recovering the cyclic property and waveform of the cyclostationary signals from the convolutive mixtures. In the third simulation (setting the same parameters as in the second simulation), the separability of the proposed algorithm is shown in Figure 7, where the convolutive mixtures are successfully separated from the two channels in terms of the distribution of the global parameters. By resorting to the multichannel row intersymbol interference (Row ISI) defined as

$$\mathrm{ISI}(i) = \frac{\sum_i \sum_k |c_{ij}(k)|^2 - \max_{j,k}|c_{ij}(k)|^2}{\max_{j,k}|c_{ij}(k)|^2} \tag{33}$$

where c_{ij} are the filter elements of the global system $\mathbf{C}(z,k)$, we can further measure the convergence speed of the proposed algorithm. From Figure 8, we see that the proposed algorithm needs a considerably smaller number of iterations to converge to the steady-state value, e.g. for ISI(1) to reach 0.01, there is approximately 30% reduction in convergence time, thereby further indicating an improved convergence performance [15].

5.2.2. Real recordings. Biomedical data frequently originate from periodic phenomena, such as breathing, tremor or contraction of the cardiac muscle [31], and since cyclostationary signals, man-made or otherwise, arise when the underlying process generating them has oscillatory

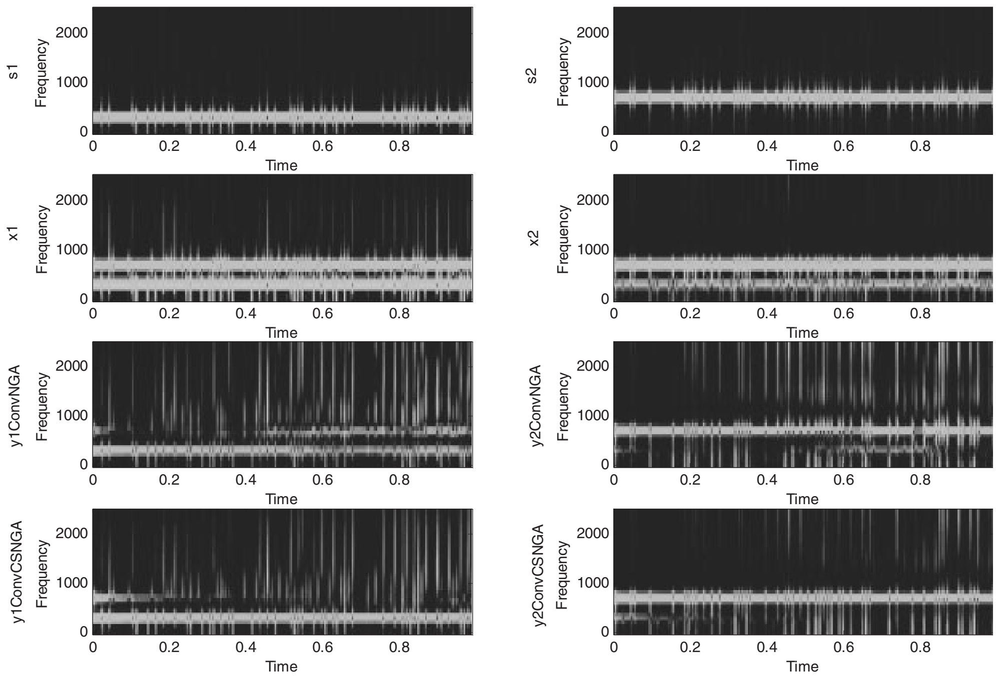

Figure 6. Separation of two BPSK source signals; the meanings of the symbols are identical to those in Figure 5.

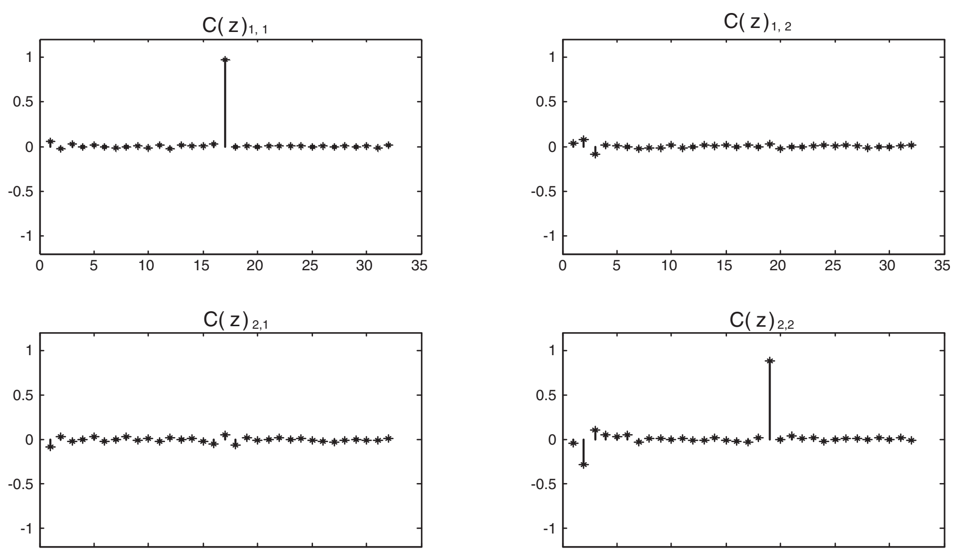

Figure 7. Simulation results of the global parameters of $\mathbf{C}(z)$ of the TITO system after convergence.

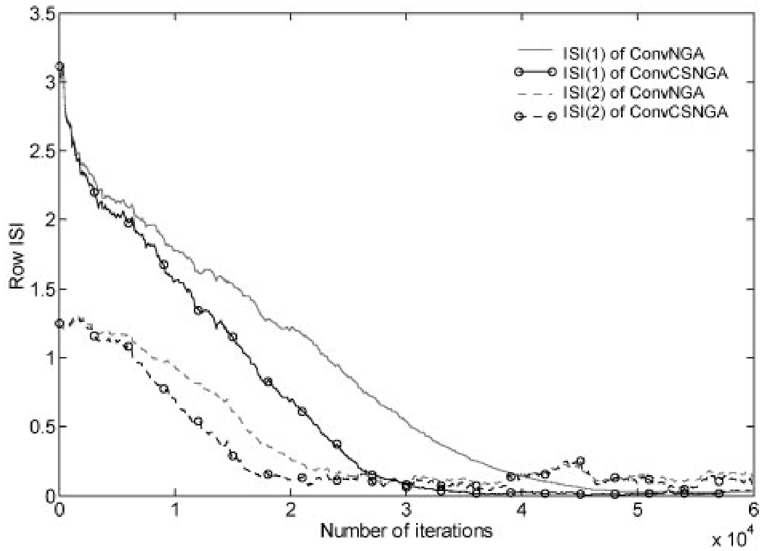

Figure 8. Comparison of convergence speed of the proposed cyclostationary convolutive NGA (ConvCSNGA) and convolutive NGA (ConvNGA) algorithm based on Row ISI.

behaviour, medical data often posses, at least, approximate cyclostationary statistical properties. ECG recordings belong to this class of signals [32, 24, 25], as their statistics repeat periodically with the occurrence of each QRS complex [33]. Therefore, addressing the problem of foetal ECG extraction using the ConvCSNGA approach has the advantage of exploiting available information about the original source signals that are sought to be separated, and which is not taken into consideration by the conventional natural gradient algorithm. Figure 9 shows the ECG signals taken from an expectant mother, recorded over 2 s at a sampling frequency of 500 Hz. The upper five recordings $x1$–$x5$ were obtained from electrodes placed on the abdominal area, and represent mixtures of both maternal and foetal contributions, as well as noise, as can be clearly seen from the waveforms. The lower three signals $x6$–$x8$ were recorded from electrodes located in the thoracic region. Clearly, no foetal contributions are visible in this case, due to the large amplitude of the maternal heart beat, and the distance between the sensors and the foetus. The components extracted by ConvNGA and ConvCSNGA are shown, respectively, in Figures 11 and 12, where all of the components are normalized to unit energy and only 500 samples are plotted so that the waveforms are clearly shown. Here, we use the cycle frequency determinator in Figure 1 to estimate the cycle frequencies that are required for the proposed algorithm, and the parameters for this determinator are stated in Section 4. As an example, the estimated cycle frequencies from the first and the last recovered ECG signals with this determinator are plotted in Figure 10. Besides the maternal and foetal heart components, the electrocardiogram generally contains several other contributions, some of which are oscillatory, including respiratory motion, tremor artifacts and periodic power line noise. From Figure 10, it is clear to see such effects by observing the difference of the cycle frequencies between various recovered signals. Therefore, it is natural to exploit the estimated cycle frequencies of each output in the iterative learning. In the simulation, the cycle frequencies are

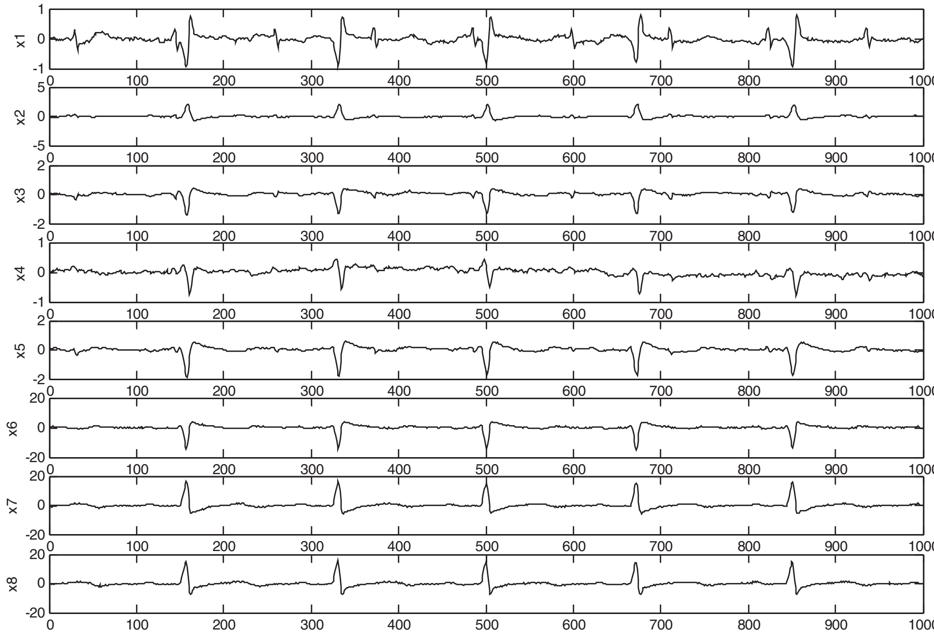

Figure 9. ECG mixtures obtained from an expectant mother. The signals $x1$–$x5$ were recorded from electrodes located in the abdominal region, while the signals $x6$–$x8$ originate from sensors in the thoracic region.

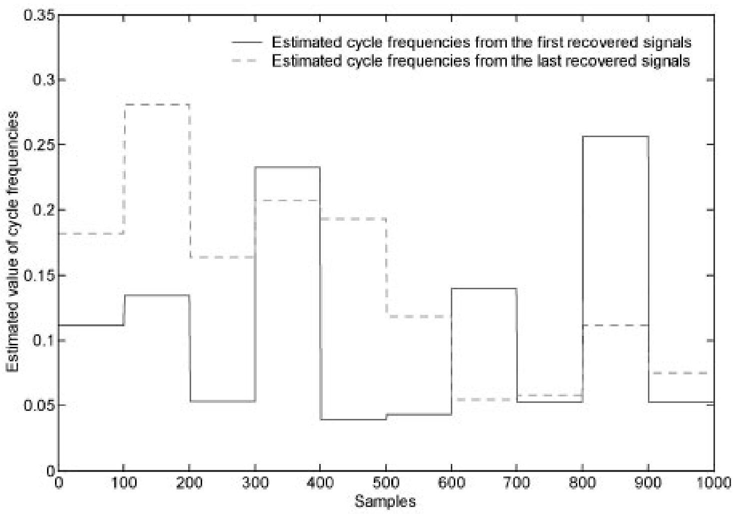

Figure 10. Estimated cycle frequencies with the proposed cycle frequency determinator.

sequentially estimated from each recovered signal with block length $B = 100$ and are then all fed into the ConvCSNGA unit in Figure 1 in the next iteration. The parameters μ, ρ and λ are set to, respectively, 0.0085, 10^{-7} and 0.03 and the filter length of the separation matrix is set to be

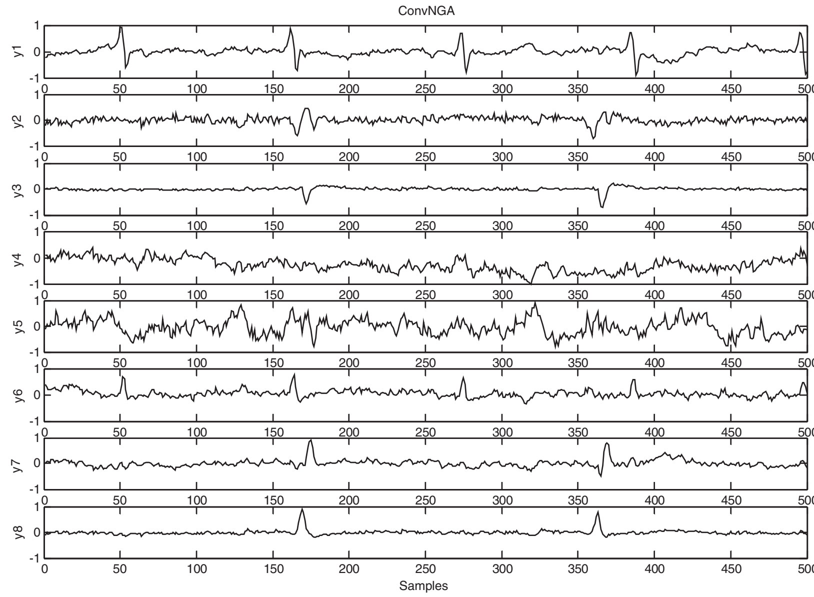

Figure 11. Separation results with ConvNGA algorithm for ECG signals in Figure 9.

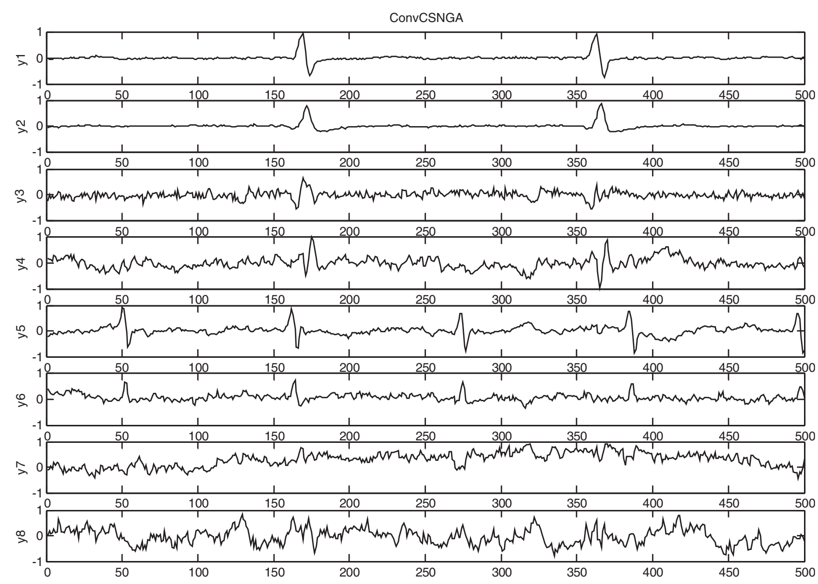

Figure 12. Separation results with ConvCSNGA algorithm for the ECG signals in Figure 9.

$L = 32$. Since the source signals are real valued, the proposed algorithm for real value is used in this case. Observing the recovered signals in Figure 12, we see that four maternal components and two foetal heart components are successfully extracted from the mixtures, and the remaining two signals may be the other interferences and noise components, which is similar to that with ConvNGA algorithm (see Figure 11). Comparing $y1$ and $y2$ in Figure 12 with the corresponding components $y3$ and $y8$ in Figure 11, it can be seen that the foetal contributions of the ECG signals recovered by the ConvCSNGA algorithm have improved waveforms which are stronger and smoother with ConvCSNGA than those with the ConvNGA algorithm. As discussed in Reference [18], the robustness of the proposed algorithm for real recordings highly relies on the estimated cycle frequencies, and using a set of cycle frequencies from one output usually reveals a different performance from the other output. Exploiting the whole cycle frequencies from all outputs reveals a better performance as compared with exploiting only one output.

6. CONCLUSIONS

A blind source separation algorithm for separating convolutive mixtures of cyclostationary signals has been presented. Exploiting the statistical cyclostationarity of signals with the natural gradient algorithm for convolutive mixtures leads to a new member (ConvCSNGA) of the family of natural gradient algorithms in adaptive source separation. One of the important implementation problems, i.e. cycle frequency determination for each estimated source signals is detailed and an efficient sequential algorithm for cycle frequency estimation based on HIF is proposed. Simulation results have shown that exploiting the cyclostationarity of signals in the blind source separation algorithm leads to faster speed of convergence, together with a better performance for the separation of the convolved cyclostationary signals, in particular in forms of shape preservation, as compared to Amari's conventional natural gradient algorithm for convolutive mixtures.

APPENDIX A

Derivation of Equation (11) (*see* Reference [18]): In terms of $\mathbf{y}(k) = \mathbf{W}(k)\mathbf{x}(k)$, the output cyclic correlation matrix is given by

$$\tilde{\mathbf{R}}_{\mathbf{y}}^{\beta}(k) = \sum_{i=1}^{N} \langle e^{J\beta_i k}\mathbf{y}(k)\mathbf{y}^{\mathrm{T}}(k)\rangle = \sum_{i=1}^{N} \mathbf{W}(k)\mathbf{R}_{\mathbf{x}}^{\beta_i}(k)\mathbf{W}^{\mathrm{T}}(k)$$

$$= \mathbf{W}(k)\left(\sum_{i=1}^{N} \mathbf{R}_{\mathbf{x}}^{\beta_i}(k)\right)\mathbf{W}^{\mathrm{T}}(k) = \mathbf{W}(k)(\tilde{\mathbf{R}}_{\mathbf{x}}^{\beta}(k))\mathbf{W}^{\mathrm{T}}(k) \tag{A1}$$

where $\tilde{\mathbf{R}}_{\mathbf{x}}^{\beta}(k) = \sum_{i=1}^{N} \mathbf{R}_{\mathbf{x}}^{\beta_i}(k)$. Applying the stochastic gradient descent, the gradient of (8) is

$$\frac{\partial \rho(\mathbf{W}(k))}{\partial \mathbf{W}(k)} = \mathbf{f}(\mathbf{y}(k))\mathbf{x}^{\mathrm{T}}(k) - (\mathbf{W}^{-1}(k))^{\mathrm{T}} + \mathbf{W}(k)\tilde{\mathbf{R}}_{\mathbf{x}}^{\beta}(k) - (\mathbf{W}^{-1}(k))^{\mathrm{T}} \tag{A2}$$

and the update equation of (11) can be easily obtained by post-multiplying $\mathbf{W}^{\mathrm{H}}(k)\mathbf{W}(k)$ to the two sides of Equation (A2).

Derivation of Equations (14) and (15):

$$
\begin{aligned}
\mathrm{d}\{\log\det(\tilde{\mathbf{R}}_{\mathbf{y}}^{\boldsymbol{\beta}}(k))\} &= \mathrm{d}\left\{\log\det\left[\sum_{i=1}^{N}\langle \mathrm{e}^{\mathrm{J}\beta_i k}\mathbf{y}(k)\mathbf{y}^{\mathrm{T}}(k)\rangle\right]\right\} \\
&\simeq \mathrm{d}\left\{\log\det\left[\left(\sum_{i=1}^{N}\mathrm{e}^{\mathrm{J}\beta_i k}\right)\langle \mathbf{y}(k)\mathbf{y}^{\mathrm{T}}(k)\rangle\right]\right\} \\
&= \mathrm{d}\left\{N\log\left(\sum_{i=1}^{N}\mathrm{e}^{\mathrm{J}\beta_i k}\right)\right\} + \mathrm{d}\{\log\det\langle \mathbf{y}(k)\mathbf{y}^{\mathrm{T}}(k)\rangle\} \\
&= \mathrm{d}\left\{N\log\left(\sum_{i=1}^{N}\mathrm{e}^{\mathrm{J}\beta_i k}\right)\right\} + \mathrm{d}\{\log\det\tilde{\mathbf{R}}_{\mathbf{x}}\} \\
&\quad + 2\,\mathrm{Tr}\{\mathrm{d}\mathbf{W}(z,k)\mathbf{W}^{-1}(z,k)\}
\end{aligned} \tag{A3}
$$

$$
\begin{aligned}
\mathrm{d}\{\mathrm{Tr}(\tilde{\mathbf{R}}_{\mathbf{y}}^{\boldsymbol{\beta}}(k))\} &= \mathrm{d}\left\{\mathrm{Tr}\left(\left(\sum_{i=1}^{N}\mathrm{e}^{\mathrm{J}\beta_i k}\right)\langle \mathbf{y}(k)\mathbf{y}^{\mathrm{T}}(k)\rangle\right)\right\} \\
&\simeq \left(\sum_{i=1}^{N}\mathrm{e}^{\mathrm{J}\beta_i k}\right)\mathrm{d}\left\{\sum_{i=1}^{N}\langle y_i^2(k)\rangle\right\} \\
&= 2\left(\sum_{i=1}^{N}\mathrm{e}^{\mathrm{J}\beta_i k}\right)\sum_{i=1}^{N}\langle y_i(k)\,\mathrm{d}y_i(k)\rangle \\
&= 2\left(\sum_{i=1}^{N}\mathrm{e}^{\mathrm{J}\beta_i k}\right)\langle \mathbf{y}^{\mathrm{T}}(k)\,\mathrm{d}\mathbf{y}(k)\rangle \\
&= 2\left(\sum_{i=1}^{N}\mathrm{e}^{\mathrm{J}\beta_i k}\right)\langle \mathbf{y}^{\mathrm{T}}(k)\,\mathrm{d}\mathbf{X}(z,k)\mathbf{y}(k)\rangle
\end{aligned} \tag{A4}
$$

ACKNOWLEDGEMENT

This work was supported by the Engineering and Physical Sciences Research Council (EPSRC) of the U.K.

REFERENCES

1. Haykin S. *Unsupervised Adaptive Filtering*. Wiley: New York, 2000.
2. Wang W, Chambers JA, Sanei S. A joint diagonalization method for convolutive blind separation of nonstationary sources in the frequency domain. In *Proceedings of the ICA2003*, 939–944, Nara, Japan, April 1–4, 2003.
3. Amari S, Douglas S, Cichocki A, Yang H. Novel on-line algorithms for blind deconvolution using natural gradient approach. *Proceedings of the SYSID-97*, Kitakyushu, Japan, July, 1997; 1057–1062.
4. Smaragdis P. Blind separation of convolved mixtures in the frequency domain. *Neurocomputing* 1998; **22**:21–34.
5. Parra L, Spence C. Convolutive blind source separation of nonstationary sources. *IEEE Transactions on Speech and Audio Processing* 2000; **8**(5):320–327.
6. Mansour A, Jutten C, Loubaton P. Adaptive subspace algorithm for blind separation of independent sources in convolutive mixtures. *IEEE Transactions on Signal Processing* 2000; **48**(2):583–586.

7. Meraim KA, Xiang Y, Manton JH, Hua Y. Blind source separation using second-order cyclostationary statistics. *IEEE Transactions on Signal Processing* 2001; **49**(4):694–701.
8. Ferreol A, Chevalier P. On the behavior of current second and higher order blind source separation methods for cyclostationary sources. *IEEE Transactions on Signal Processing* 2000; **48**(6):1712–1725.
9. Jafari MG, Chambers JA, Mandic DP. Natural gradient algorithm for cyclostationary sources. *IEE Electronic Letters* 2002; **38**(7):758–759.
10. Gardner WA. *Statistical Spectral Analysis: A Nonprobabilistic Theory*. Prentice-Hall: Englewood Cliffs, NJ; 1988.
11. Jafari MG, Alty S, Chambers JA. A new natural gradient algorithm for cyclostationary sources. *IEE Proceedings—Vision Image and Signal Processing* 2004; **151**(1).
12. Cichocki A, Amari S. *Adaptive Blind Signal and Image Processing: Learning Algorithms and Applications*. Wiley: New York, 2002.
13. Cardoso JF, Laheld B. Equivariant adaptive source separation. *IEEE Transactions on Signal Processing* 1996; **44**(12):3017–3030.
14. Choi S, Cichocki A, Amari S. Equivariant nonstationary source separation. *Neural Networks* 2002; **15**(1):121–130.
15. Wang W, Jafari MG, Sanei S, Chambers JA. Blind separation of convolutive mixtures of cyclostationary sources using an extended natural gradient method. In *Proceedings of the ISSPA2003*, II: 93–96, Paris, France, July 1–4, 2003.
16. Zhang L, Cichocki A, Amari S. Geometrical structures of FIR manifold and their application to multichannel blind deconvolution. *Journal of VLSI for Signal Processing* 2002; **31** 31–44.
17. Amari S, Chen TP, Cichocki A. Stability analysis of learning algorithms for blind source separation. *Neural Networks* 1997; **10**(8):1345–1351.
18. Jafari MG. Novel sequential algorithms for blind source separation of instantaneous mixtures. *Ph.D. Thesis* of King's College, University of London, 2002.
19. Barros AK, Ohnishi N. Heart instantaneous frequency (HIF): an alternative approach to extract heart rate variability. *IEEE Transactions on Biomedical Engineering* 2001; **48**(7):850–855.
20. Cichocki A, Amari S, Siwek K, *et al*. ICALAB Toolboxes. *http://www.bsp.brain.riken.go.jp*.
21. Tong L, Soon V, Huang YF, Liu R. Indeterminacy and identifiability of blind identification. *IEEE Transactions on CAS of VLSI for Signal Processing* 1991; **38**:499–509.
22. Tong L, Inouye Y, Liu R. Waveform-preserving blind estimation of multiple independent sources. *IEEE Transactions on Signal Processing* 1993; **41**(7):2461–2470.
23. Belouchrani A, Abed-Meraim K, Cardoso JF, Moulines E. A blind source separation technique using second order statistics. *IEEE Transactions on Signal Processing* 1997; **45**(2):434–444.
24. De Lathauwer L, De Moor B, Vandewalle J. Fetal electrocardiogram extraction by blind source subspace separation. *IEEE Transactions on Biomedical Engineering*—Special Topic Section on Advances in Statistical Signal Processing for Medicine 2000; **47**(5):567–572.
25. Zarzoso V, Nandi AK. Noninvasive fetal electrocardiogram extraction: blind separation versus adaptive noise cancellation. *IEEE Transactions on Biomedical Engineering* 2001; **48**(1):12–18.
26. Cardoso JF, Souloumiac A. Jacobi angles for simultaneous diagonalization. *SIAM Journal of Matrix Analysis and Applications* 1996; **17**(1) 161–164.
27. Georgiev P, Cichocki A. Robust blind source separation utilizing second and fourth order statistics. In *International Conference on Artificial Neural Networks (ICANN2002)* 2002; Mardrid, Spain, August, 2002.
28. Hyvarinen A, Oja E. A fast fixed-point algorithm for independent component analysis. *Neural Computation* 1997; **9**(7):1483–1492.
29. Amari S, Cichocki A, Yang HH. A new learning algorithm for blind signal separation. *Advances in Neural Information Processing Systems, NIPS-1995*, vol. 8 MIT Press: Cambridge, MA, 1996; 757–763.
30. Amari S, Cichocki A. Adaptive blind signal processing—neural network approaches. *Proceedings IEEE* 1998; **86**:1186–1187.
31. Timmer J. Power of surrogate data testing with respect to nonstationarity. *Physical Review E* 1998; **58**:5153–5156.
32. Saha S, Ramakrishnan AG. Transmission of chosen transform coefficients of normalized cardiac beats for compression. *Proceedings of the ICASSP*, 1997; vol. **3**, 1901–1904.
33. Speirs CA, Soraghan JJ, Stewart RW, Byrne S. Least squares time sequenced adaptive filtering for the detection of fragmented micropotentials. *http://www.spd.eee.strath.ac.uk/research.html*, 1995.

