

# BLIND SEPARATION OF CONVOLUTIVE MIXTURES OF CYCLOSTATIONARY SOURCES USING AN EXTENDED NATURAL GRADIENT METHOD

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

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

E-mail: [wenwu.wang, maria.jafari, saeid.sanei, jonathon.chambers]@kcl.ac.uk

## ABSTRACT

An on-line adaptive blind source separation algorithm for the separation of convolutive mixtures of cyclostationary source signals is proposed. The algorithm is derived by applying natural gradient iterative learning to the novel cost function which is defined according to the wide sense cyclostationarity of signals. 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.

## 1. INTRODUCTION

In many practical situations such as in the radiocommunications, telemetry, radar, sonar and speech contexts, the sources are nonstationary and often (quasi)-cyclostationary and the observed signals are usually convolutive mixtures, so that the conventional methods for 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, are not appropriate any more. Increasing interest has therefore been focused on solving the problem of BSS of convolutive mixtures [1].

Addressing the BSS problem for cyclostationary sources is a relatively new approach. In [2], the proposed method minimizes a cost function constructed from the cyclic cross-correlation of recovered sources at various time lags, and presents iterative update equations following from the natural gradient technique. In [3], it is 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 [3] or are based on second-order statistics as in [2], and very few of them are concerned about convolutive mixtures. In this paper, we propose an on-line adaptive

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

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 of the convolutive mixtures of cyclostationary sources.

## 2. PROBLEM FORMULATION

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

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

where  $\mathbf{s}(k) \in \mathbb{C}^N$  is the source vector,  $\mathbf{x}(k) \in \mathbb{C}^M$  is the sensor vector,  $k$  is the discrete time index;  $\mathbf{H}(z) \in \mathbb{C}^{M \times N}$  is the  $z$ -transform of the mixing matrix with entries  $H_{ij}(z) = \sum_{p=0}^{P-1} h_{ijp} z^{-p}$  ( $i = 1, \dots, M; j = 1, \dots, N$ ), 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 modeled as a wide sense cyclostationary signal [4], 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

$$\begin{aligned} \mathbf{y}(k) &= \mathbf{W}(z, k)\mathbf{x}(k) \\ &= \mathbf{W}(z, k)\mathbf{H}(z)\mathbf{s}(k) = \mathbf{C}(z, k)\mathbf{s}(k) \end{aligned} \quad (2)$$

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  $W_{ij}(z, k) = \sum_{p=0}^P w_{ijp}(k)z^{-p}$ . The task is to adjust  $\mathbf{W}(z, k)$  such that  $\lim_{k \rightarrow \infty} \mathbf{C}(z, k) = \mathbf{P}\mathbf{A}\mathbf{D}(z)$ , where  $\mathbf{P} \in \mathbb{R}^{N \times N}$  is a permutation matrix,  $\mathbf{A} \in \mathbb{R}^{N \times N}$  is a nonsingular diagonal scaling matrix, and  $\mathbf{D}(z)$  is a diagonal matrix whose  $ii$ -th 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 [4], we assume that the cyclic correlation matrix of the sources at lag  $\tau = 0$  follows  $\mathbf{R}_s^{\beta_i}(0) = E\{e^{J\beta_i k} \mathbf{s}(k) \mathbf{s}^H(k)\} = \mathbf{I}$ , where  $(\cdot)^H$  denotes the Hermitian transpose operator, and  $\mathbf{I}$  has entries as follows

$$[\mathbf{I}]_{l,g} = \begin{cases} 1, & \text{if } l \in \{1, 2, \dots, N\}, g = l = i \\ 0, & \text{otherwise} \end{cases} \quad (4)$$

### 3. THE PROPOSED ALGORITHM

#### 3.1. Combining Cyclic Decorrelation With Natural Gradient Learning for Instantaneous Mixtures

The wide sense cyclostationarity assumption implies that the cyclic correlation function of the sources satisfies [2]:

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

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

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

where  $J = \sqrt{-1}$ , the superscript  $*$  denotes complex conjugation,  $\langle \cdot \rangle$  denotes the time averaging operator, and  $\beta_i$  is a nonzero cycle frequency of source  $i$ . Invoking these properties into the Kullback-Leibler divergence, our cost function for instantaneous mixtures is defined as

$$\begin{aligned} \rho(\mathbf{W}(k)) &= -\log \det(\mathbf{W}(k)) - \sum_{i=1}^N \log p_i(y_i(k)) \\ &+ \frac{1}{2} \{Tr(\tilde{\mathbf{R}}_{yy}^\beta(k)) - \log \det(\tilde{\mathbf{R}}_{yy}^\beta(k)) - N\} \end{aligned} \quad (8)$$

where  $Tr(\cdot)$  and  $\det(\cdot)$  are respectively the trace and determinant operators, and  $p_i(y_i(k))$  is an appropriately chosen independent pdf. The term  $\tilde{\mathbf{R}}_{yy}^\beta(k)$  is defined as  $\tilde{\mathbf{R}}_{yy}^\beta(k) = \sum_{i=1}^N \mathbf{R}_{yy}^{\beta_i}(k)$ , where  $\mathbf{R}_{yy}^{\beta_i}(k) = \langle e^{J\beta_i k} \mathbf{y}(k) \mathbf{y}^H(k) \rangle$  is the output cyclic correlation matrix for the  $i$ -th cycle frequency which is required to satisfy  $\lim_{k \rightarrow \infty} \tilde{\mathbf{R}}_{yy}^{\beta_i}(k) = \mathbf{I}$ , where the elements of  $\mathbf{I}$  take the form in (4). At convergence,  $\lim_{k \rightarrow \infty} \tilde{\mathbf{R}}_{yy}^\beta(k) = \mathbf{I}$ . It is straightforward to follow that iterative learning will result in the minimization of this cost function. Applying the natural gradient rule [5] together with a cyclic decorrelation operation, we obtain a new learning rule as

$$\begin{aligned} \mathbf{W}(k+1) &= \mathbf{W}(k) + \mu(k) [\mathbf{I} - \mathbf{f}(\mathbf{y}(k)) \mathbf{y}^H(k) \\ &+ \mathbf{I} - \tilde{\mathbf{R}}_{yy}^\beta(k)] \mathbf{W}(k) \end{aligned} \quad (9)$$

where  $\mu(k)$  is the learning rate. It is necessary to use split-complex nonlinearities in the complex case, e.g.  $f_i(y_i(k)) =$

$\tanh(y_{iR}(k)) + J \tanh(y_{iI}(k))$ , where  $y_{iR}(k)$  and  $y_{iI}(k)$  are respectively the real and imaginary parts of  $y_i(k)$ , for the sources with super-Gaussian distributions, or  $f_i(y_i(k)) = y_i(k) |y_i|^2$  for the sub-Gaussian case.

#### 3.2. On-line Adaptive Learning for Separating the Convolutional Mixtures of Cyclostationary Signals

For convolutive BSS, from the algebraic point of view, they have equivalent mathematical models to the instantaneous cases, only differing in the way in which (IIR filter, FIR filter,  $\mathcal{Z}$ -transform, wavelets, other transforms) the physical phenomena are described [5]. Exploiting this fact, it is straightforward to use the  $\mathcal{Z}$ -transform and define the following cost function for convolutive mixtures of cyclostationary sources

$$\rho(\mathbf{W}(z, k)) = \rho_\gamma(\mathbf{W}(z, k)) + \rho_\beta(\mathbf{W}(z, k)) \quad (10)$$

where the quantity  $\rho_\gamma(\mathbf{W}(z, k))$  is given by  $\rho_\gamma(\mathbf{W}(z, k)) = -\sum_{i=1}^N \log p_i(y_i(k)) - \frac{1}{2\pi J} \oint \log |\det \mathbf{W}(z, k)| z^{-1} dz$ ,  $\rho_\beta(\mathbf{W}(z, k)) = \frac{1}{2} \{Tr(\tilde{\mathbf{R}}_{yy}^\beta(k)) - \log \det(\tilde{\mathbf{R}}_{yy}^\beta(k)) - N\}$ . 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 (10), we follow the derivation of the algorithm presented by Amari *et al.* [5] 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 on the manifold of FIR filters  $\mathcal{M}(L)$ , i.e.  $d\mathbf{X}(z, k) = d\mathbf{W}(z, k) \otimes \mathbf{W}^\dagger(z)$ , where  $\otimes$  and  $^\dagger$  are the multiplication and inverse operation defined in the manifold  $\mathcal{M}(L)$ ,  $d\rho_\gamma(\mathbf{W}(z, k))$  takes the following form by simple algebraic and differential calculus

$$d\rho_\gamma(\mathbf{W}(z, k)) = \mathbf{f}^H(\mathbf{y}(k)) d\mathbf{X}(z, k) [\mathbf{y}(k)] - Tr(d\mathbf{X}(z, k)) \quad (11)$$

Given that  $\mathbf{y}(k) = \mathbf{W}(z, k) \mathbf{x}(k)$ ,  $\tilde{\mathbf{R}}_{xx} = \langle \mathbf{x}(k) \mathbf{x}^T(k) \rangle$ ,  $d\mathbf{y}(k) = d\mathbf{X}(z, k) \mathbf{y}(k)$ , and the truncated form of  $d\mathbf{X}(z, k)$ , i.e.  $[d\mathbf{W}(z, k) \mathbf{W}^{-1}(z, k)]_L$ , we have

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

$$\begin{aligned} d\{\log \det(\tilde{\mathbf{R}}_{yy}^\beta(k))\} &= d\{N \log(\sum_{i=1}^N e^{J\beta_i k})\} \\ &+ d\{\log \det \tilde{\mathbf{R}}_{xx}\} + 2Tr\{d\mathbf{W}(z, k) \mathbf{W}^{-1}(z, k)\} \end{aligned}$$

where  $d\mathbf{X}(z, k)$  is denoted as  $\sum_{p=0}^L d\mathbf{X}_p(k) z^{-p}$ . Notice that  $d\{N \log(\sum_{i=1}^N e^{J\beta_i k})\}$  and  $d\{\log \det \tilde{\mathbf{R}}_{xx}\}$  do not depend on the weight matrix  $\mathbf{W}(z, k)$ ,  $d\{\log \det(\tilde{\mathbf{R}}_{yy}^\beta(k))\}$

can be reduced to  $2\text{tr}\{d\mathbf{W}(z,k)\mathbf{W}^{-1}(z,k)\}$ , therefore we have ( $N$  does not depend on  $\mathbf{W}(z,k)$  either)

$$\begin{aligned} d\rho_{\beta}(\mathbf{W}(z,k) &= -\text{Tr}\{d\mathbf{W}(z,k)\mathbf{W}^{-1}(z,k)\} \\ &+(\sum_{i=1}^N e^{j\beta_i k} \langle \mathbf{y}^H(k) d\mathbf{X}(z,k) \mathbf{y}(k) \rangle) \end{aligned} \quad (12)$$

From (11) and (12), the partial derivation of  $\rho(\mathbf{W}(z,k))$  with respect to  $\mathbf{X}(z,k)$  leads to

$$\begin{aligned} \frac{\partial \rho(\mathbf{W}(z,k))}{\partial \mathbf{X}_p(k)} &= \{\mathbf{f}(\mathbf{y}(k))\mathbf{y}^H(k-p) - \mathbf{I}\delta_p\} \\ &+ \{-\mathbf{I}\delta_p + \sum_{i=1}^N \langle e^{j\beta_i k} \mathbf{y}(k) \mathbf{y}^H(k-p) \rangle\} \end{aligned} \quad (13)$$

The truncated form of  $d\mathbf{X}(z,k)$  on the FIR filter manifold  $\mathcal{M}(L)$ , i.e.  $[d\mathbf{W}(z,k)\mathbf{W}^{-1}(z,k)]_L$ , implies that the partial derivation of  $\rho(\mathbf{W}(z,k))$  with respect to  $\mathbf{W}(z,k)$  can be obtained according to the partial derivation with respect to  $\mathbf{X}(z,k)$ , applying the moving average approach to the last term of (13), that is,  $\hat{\mathbf{R}}_{\mathbf{y}\mathbf{y}}^{\beta_i(k)}(p) = (1 - \eta_0)\hat{\mathbf{R}}_{\mathbf{y}\mathbf{y}}^{\beta_i(k)}(p) + \eta_0 \sum_{i=1}^N e^{j\beta_i k} \mathbf{y}(k) \mathbf{y}^H(k-p)$ , an on-line learning algorithm for convolutive mixtures of cyclostationary signals can be implemented as

$$\begin{aligned} \mathbf{W}_p(k+1) &= \mathbf{W}_p(k) + \mu(k)\{\mathbf{W}_p(k) - \hat{\mathbf{Q}}_q(k) \\ &+ \mathbf{W}_p(k) - \mathbf{f}(\mathbf{y}(k))\mathbf{u}_{\gamma}^H(k)\} \end{aligned} \quad (14)$$

where  $\mathbf{u}_{\gamma}(k)$  can be denoted as  $\sum_{q=0}^L \mathbf{W}_{L-q}^H(k)\mathbf{y}(k-q)$ , and  $\hat{\mathbf{Q}}_q(k) = \sum_{q=0}^L \mathbf{W}_{L-q}^H(k)\hat{\mathbf{R}}_{\mathbf{y}\mathbf{y}}^{\beta_i(k)}(q)$ . The equilibrium points of the proposed learning algorithm satisfy that  $E\{\sum_{i=1}^N e^{j\beta_i k} \mathbf{y}(k) \mathbf{y}^H(k)\} = \mathbf{I}$ ,  $E\{\sum_{i=1}^N e^{j\beta_i k} \mathbf{y}(k) \mathbf{y}^H(k-p)\} = \mathbf{0}$ ,  $E\{\mathbf{f}(\mathbf{y}(k))\mathbf{y}^H(k-p)\} = \mathbf{0}$ ,  $E\{\mathbf{f}(\mathbf{y}(k))\mathbf{y}^H(k)\} = \mathbf{I}$ , where  $p = 1, 2, \dots, L$ . 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 avoided by following the nonholonomic constraint [5], that is, using  $\mathbf{W}_p(k)\mathbf{\Lambda}(k)$  instead of  $\mathbf{W}_p(k)$  in (14), where  $\mathbf{\Lambda}(k) = \text{diag}\{\mathbf{f}(\mathbf{y}(k))\mathbf{y}^H(k)\}$ . The detailed theoretical analysis of the convergence behavior of the proposed algorithm in the Lie group sense is given in [6].

In implementation,  $\mathbf{y}(k)$  is given by FIR filtered version of the input data, i.e.  $\mathbf{y}(k) = \sum_{p=0}^L \mathbf{W}_p(k)\mathbf{x}(k-p)$  and the general updating equation in which all signals and coefficients are complex-valued is

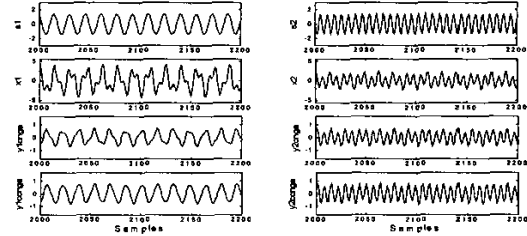
$$\begin{aligned} \mathbf{W}_p(k+1) &= \mathbf{W}_p(k) + \mu(k)\{\mathbf{W}_p(k) - \hat{\mathbf{Q}}_{p+q}(k-L) \\ &+ \mathbf{W}_p(k) - \mathbf{f}(\mathbf{y}(k-L))\mathbf{u}_{\gamma}^H(k-p)\} \end{aligned} \quad (15)$$

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  $\rho$  is a positive value which avoids an explosive growth of the step size.

#### 4. SIMULATIONS

In this section, we examine the performance of the proposed algorithm by simulation. A TITO (Two Input Two Output) system is considered, that is,  $N = M = 2$  (The algorithm is also suitable for generic MIMO systems). The resemblance between the original and the reconstructed source waveforms is measured by their mean squared difference  $\epsilon^2(dB) = 10 \log\{1 - (1/2N) \sum_{i=1}^N E[|y_i(k) - s_i(k)|^2]\}$  (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



**Fig. 1.** Separation of convolutive mixtures of two sinusoid signals:  $s_1$  and  $s_2$  are two sources,  $x_1$  and  $x_2$  are the two convolutive mixtures,  $y_1cnga$  and  $y_2cnga$  are the recovered signals by the natural gradient algorithm,  $y_1cnqa$  and  $y_2cnqa$  are the recovered signals by the proposed method.

signals are set respectively to be  $s_1(i) = \sin(0.4 * i)$ ,  $s_2(i) = \sin(0.9 * i)$ . The two sources are convolved by FIR filters with entries  $H(1, : , 1) = [1 \ 1 \ -0.75 \ 0.9]$ ,  $H(2, : , 1) = [0.5 \ -0.3 \ 0.2 \ 0.2]$ ,  $H(1, : , 2) = [-0.2 \ 0.4 \ 0.7 \ 0.2]$ ,  $H(2, : , 2) = [0.2 \ 1 \ 0 \ 0.9]$ . The other parameters are  $L = 32$ ,  $\mu = 0.004$ , and  $\rho = 0.1$ . The two cyclic frequencies are respectively  $(5\pi)^{-1}$  and  $9(20\pi)^{-1}$ . The separation results are shown in Figure 1, and the performance indices are respectively 19.67dB and 19.12dB. In the second simulation (see Figure 2), 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 19.93dB and 19.14dB. Both simulation results indicate that the proposed algorithm has

improved performance over the natural gradient 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 3, 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  $ISI(i) = \frac{\sum_i \sum_k |c_{ij}(k)|^2 - \max_{j,k} |c_{ij}(k)|^2}{\max_{j,k} |c_{ij}(k)|^2}$ , 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 4, 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. These are only preliminary simulation results which are being extended for further comparison in [6].

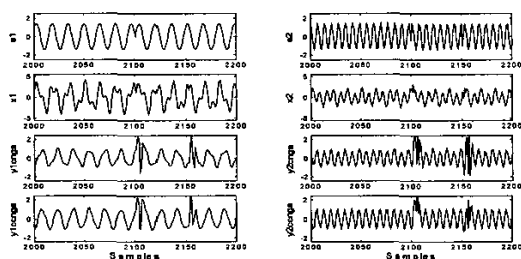


Fig. 2. Separation of two BPSK source signals; the means of the symbols are identical to those in Fig. 1.

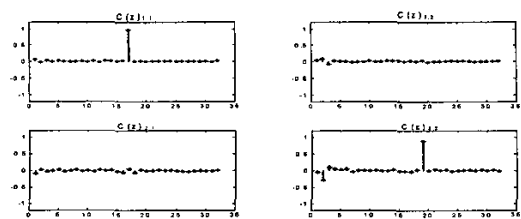


Fig. 3. Simulation results of the global parameters of  $\mathbf{C}(z)$  of the TITO system after convergence.

## 5. CONCLUSIONS

An on-line adaptive blind separation algorithm for separating convolutive mixtures of cyclostationary source signals has been presented. Simulation results have shown

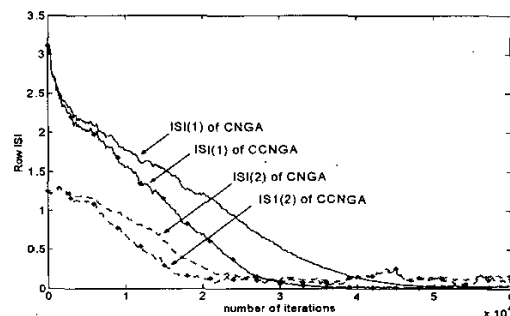


Fig. 4. Comparison of convergence speed of the proposed cyclostationary convolutive NGA (CCNGA) and convolutive NGA (CNGA) algorithm based on Row ISI.

that the 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. Algorithm implementation to more practical environment and wider kinds of signals (i.e. EEG signals) is our future work.

## 6. REFERENCES

- [1] W. Wang, J. A. Chambers, and S. Sanei, "A joint diagonalization method for convolutive blind separation of nonstationary sources in the frequency domain," *Proc. of ICA2003, Nara, Japan, April 1-4, 2003*.
- [2] K. A. Meraim, Y. Xiang, J. H. Manton, and Y. Hua, "Blind source separation using second-order cyclostationary statistics," *IEEE Trans. on SP*, vol. 49, pp. 694–701, Apr. 2001.
- [3] A. Ferreol and P. Chevalier, "On the behavior of current second and higher order blind source separation methods for cyclostationary sources," *IEEE Trans. on SP*, vol. 48, pp. 1712–1725, Jun. 2000.
- [4] M. G. Jafari, J. A. Chambers, and D. P. Mandic, "Natural gradient algorithm for cyclostationary sources," *IEE Electronic Letters*, vol. 38, pp. 758–759, Jul. 2002.
- [5] A. Cichocki and S. Amari, "Adaptive blind signal and image processing: Learning algorithms and applications," *Wiley*, April, 2002.
- [6] W. Wang, M. G. Jafari, S. Sanei, and J. A. Chambers, "Blind source separation of convolutive mixtures of cyclostationary signals," *submitted to International Journal of Adaptive Control and Signal Processing*, February, 2003.