Blind Separation of Cyclostationary Sources Using Joint Block Approximate Diagonalization

D.T. Pham

Laboratoire Jean Kuntzmann - CNRS/INPG/UJF, BP 53, 38041 Grenoble Cedex, France
Dinh-Tuan.Pham@imag.fr

Abstract. This paper introduces an extension of an earlier method of the author for separating stationary sources, based on the joint approximated diagonalization of interspectral matrices, to the case of cyclostationary sources, to take advantage of their cyclostationarity. the proposed method is based on the joint block approximate diagonlization of cyclic interspectral density. An algorithm for this diagonalization is described. Some simulation experiments are provided, showing the good performance of the method.

1 Introduction

Blind source separation aims at recovering sources from their unknown mixtures [1]. All separation methods are based on some "non properties" of the source signals. Early methods which do not exploit the time structure of the signals would require non Gaussianity of the sources. However, by exploiting the time structure, one can separate mixtures of Gaussian sources provided that the sources are *not* independent identically distributed (iid) in time., that is one (or both) of the two "i" in "iid" is not met. If only the first "i" is not met, one has stationary correlated sources and separation can be achieved by considering the lagged covariances or inter-spectra between mixtures signals. This is the basis of most second order separation methods [2, 3, 4]. If the second "i" in "iid" is not fulfilled, one has nonstationary sources and separation methods can again be developed using only second order statistics [5, 6]. However, "nonstationarity" is a too general non property to be practical, the above works actually focus only on a particular aspect of it: They assume temporal independence (or more accurately ignore possible temporal dependency) and focus only on the variation of variance of the signal in time and assume that this variation is slow enough to be adequately estimated nonparametrically. In this paper, we consider another aspect of non stationarity: the cyclostationarity. The variance of the source is also variable in time but in an (almost) periodic manner. Further, the autocovariance between the source at different time points does not depend only on the delay as in the stationary case, but also on time as well and again in a (almost) periodic manner. Thus the "nonstationary" method in [6] may not work as this source variance can vary rapidly since the period (frequency) can be short (high). Moreover, such method ignores the lagged autocovariance of the sources, which provide important useful information for the separation. The "stationary" methods [2, 3, 4] still work in general if one takes as lagged covariances the average lagged covariances

M.E. Davies et al. (Eds.): ICA 2007, LNCS 4666, pp. 244–251, 2007.

over time. In fact the usual lagged covariance estimator when applied to cyclostationary signal actually estimates the average lagged covariance. However, such methods ignore the specificity of cyclostationary signals and thus don't benefice from it and further would fail if the sources are noncorrelated (but has variance varying periodically with time). Our method is specially designed to exploit this specificity. There have been several works on blind separation of cyclostationary sources [7, 8, 9, 10]. Our work is different in that we work with cyclic inter-spectral densities while the above works are mainly based on cyclic cross-covariances. Our work may be viewed as an extension of our earlier work for blind separation of stationary sources [4] based on the joint approximate diagonalization of a set of inter-spectral matrices. As said earlier, this method still works for cyclostationary sources, provided that their average spectra are different up to a constant factor. The present method exploits the extra information of cyclostationarity and thus yields better performance and also can be dispensed with the above restriction.

2 Cyclostationary Signals

A discrete time (possibly complex value) process $\{X(t)\}$ is said to be cyclostationary (or almost periodically correlated) if its mean function $t \mapsto \mathrm{E}[X(t)]$ and its covariance functions $t \mapsto \mathrm{cov}\{X(t+\tau), X(t)\}$ are almost periodic [11]. The definition of almost periodicity is rather technical, but here we consider only zero mean cyclostationary process with a finite "number of cycles", for which an equivalent definition is that there exists a finite subset $\mathcal{A}$ of $(-1/2, 1/2]$ such that

$$\mathrm{E}[X(t+\tau)X^*(t)] = \sum_{\alpha\in\mathcal{A}} R(\alpha;\tau)e^{i2\pi\alpha t}, \quad \forall t,\ \forall\tau. \tag{1}$$

where * denotes the complex conjugate. The function $\tau \mapsto R(\alpha;\tau)$ is called the cyclic autocovariance function of cycle α. From (1), it can be computed as

$$R(\alpha;\tau) = \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \mathrm{E}[X(t+\tau)X^*(t)]e^{-i2\pi\alpha t} \tag{2}$$

Note that for $\alpha \notin \mathcal{A}$, the last right hand side yields zero by (1). Thus we may define $R(\alpha,\tau)$ for all α, τ by the above right hand side, and $\mathcal{A}$ as the set $\{\alpha : R(\alpha;\cdot) \neq 0\}$.

We shall assume that the function $R(\alpha;\cdot)$ admits a Fourier transform $f(\alpha;\cdot)$, called the cyclic spectral density of cycle α:

$$f(\alpha;\nu) = \sum_{\tau=-\infty}^{\infty} R(\alpha;\tau)e^{-i2\pi\nu\tau} \quad \Leftrightarrow \quad R(\alpha;\tau) = \int_0^1 f(\alpha;\nu)e^{i2\pi\nu\tau}d\nu.$$

Note It can be seen from (2) that $R(-\alpha;\tau) = R^*(\alpha;-\tau)e^{-i2\pi\alpha\tau}$. This means that if $\mathcal{A}$ contains α, it must contain $-\alpha$.

Let $\alpha_1, \ldots, \alpha_q$ be in $\mathcal{A}$, the matrix of general j,k element $R(\alpha_k - \alpha_j;\tau)e^{i2\pi\alpha_j\tau}$ can be seen to be the average autocovariance of lag τ of the vector process $\{[X(t)e^{i2\pi\alpha_1 t} \ \cdots \ X(t)e^{i2\pi\alpha_q t}]^T\}$, since

$$R(\alpha_k-\alpha_j;\tau)e^{i2\pi\alpha_j\tau}=\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\mathrm{E}[X(t+\tau)X^*(t)]e^{i2\pi\alpha_j(t+\tau)}e^{-i2\pi\alpha_k t}$$

Therefore this matrix as a function of τ is of type positive and it follows that its Fourier transform is a non negative (matrix) function. In other words:

$$\begin{bmatrix} f(0;\nu-\alpha_1) & \cdots & f(\alpha_q-\alpha_1;\nu-\alpha_1)\\ \vdots & \ddots & \vdots\\ f(\alpha_1-\alpha_q;\nu-\alpha_q) & \cdots & f(0;\nu-\alpha_q)\end{bmatrix}\geq\mathbf{0} \tag{3}$$

In particular, $f(0;\cdot)\geq 0$. The functions $R(0;\cdot)$ and $f(0;\cdot)$ may be viewed as the average covariance function and spectral density of the process $\{X(t)\}$. Since $R(0;0)=\lim_{T\to\infty}T^{-1}\sum_{t=1}^{T}\mathrm{E}[|X(t)|^2]>0$, $0\in\mathcal{A}$. By taking $\alpha_1=0$, one see that the matrix in (3) can contain all the cyclic spectral densities of cycle in $\mathcal{A}$ and possibly some other vanishing cyclic spectral densities (since its cycle is not in $\mathcal{A}$) as well.

The natural estimator of $R(\alpha;\tau)$ based on an observed sample $X(1),\ldots,X(T)$ is

$$\hat{R}(\alpha;\tau)=\frac{1}{T}\sum_{t=\max(1,1-\tau)}^{\min(T,T-\tau)}X(t+\tau)X^*(t)e^{-i2\pi\alpha t}. \tag{4}$$

From this estimator, one may construct an estimator for $f(\alpha;\nu)$ as

$$\hat{f}(\alpha;\nu)=\sum_{\tau=1-T}^{T-1}k_M(\tau)\hat{R}(\alpha;\tau)e^{-i2\pi\nu\tau} \tag{5}$$

where $k_M(\cdot)$ is a given lag windows, often of the form $k(\cdot/M)$ with k being some given even function taking the value 1 at 0, and M is a window width parameter.

3 The Mixture Model and Separation Method

We consider the simplest mixture model in which the mixing is instantaneous without noise and there is a same numbers of mixtures as the sources: $\mathbf{X}(t)=\mathbf{A}\mathbf{S}(t)$ where $\mathbf{X}(t)$ and $\mathbf{S}(t)$ denote the vectors of mixtures and of sources at time t, and $\mathbf{A}$ is a square matrix. The sources are assumed to be independent cyclostationary processes. It is easily seen that the observed mixtures are also cyclostationary, with the set of cycle frequencies contained in the union of the sets of cycle frequencies of the sources, which we denote by $\mathcal{A}$. The goal is to recover the sources from their mixtures. For simplicity, we shall assume that $\mathcal{A}$ is known. In practice, such set can be estimated. Further, it is not important that $\mathcal{A}$ be accurately known.

We define the cyclic autocovariance function $\mathbf{R}_{\mathbf{X}}(\alpha;\cdot)$ of cycle α of the vector process $\{\mathbf{X}(t)\}$ similar to (2) except that $X(t)$ is replaced by $\mathbf{X}(t)$ and * is understood as the transpose conjugate. Clearly $\mathbf{R}_{\mathbf{X}}(\alpha;\tau)=\mathbf{A}\mathbf{R}_{\mathbf{S}}(\alpha;\tau)\mathbf{A}^*$ where $\mathbf{R}_{\mathbf{S}}(\alpha;\cdot)$ is the cyclic autocovariance function of cycle α of the vector source process $\{\mathbf{S}(t)\}$. The independence of the sources implies that the matrices $\mathbf{R}_{\mathbf{S}}(\alpha;\tau)$ are diagonal for all α,τ

(of course if $\alpha \notin \mathcal{A}$ this matrix vanishes and is of no interest). Similarly, we define the cyclic spectral density of cycle α of the vector process $\{\mathbf{X}(t)\}$ as the Fourier transform $\mathbf{f}_{\mathbf{X}}(\alpha;\cdot)$ of $\mathbf{R}_{\mathbf{X}}(\alpha;\cdot)$. Again, we have $\mathbf{f}_{\mathbf{X}}(\alpha;\nu) = \mathbf{A}\mathbf{f}_{\mathbf{S}}(\alpha;\nu)\mathbf{A}^*$ where $\mathbf{f}_{\mathbf{S}}(\alpha;\cdot)$ is the cyclic spectral density of cycle α of the vector process $\{\mathbf{S}(t)\}$, which is diagonal for all frequencies and all α.

The analogue of the matrix in (3) is the block matrix

$$\mathbf{C}(\nu) = \begin{bmatrix} \mathbf{C}_{11}(\nu) & \cdots & \mathbf{C}_{1K}(\nu) \\ \vdots & \ddots & \vdots \\ \mathbf{C}_{K1}(\nu) & \cdots & \mathbf{C}_{KK}(\nu) \end{bmatrix} \tag{6}$$

where

$$\mathbf{C}_{jk}(\nu) = \begin{bmatrix} f_{X_jX_k}(0;\nu-\alpha_1) & \cdots & f_{X_jX_k}(\alpha_q-\alpha_1;\nu-\alpha_1) \\ \vdots & \ddots & \vdots \\ f_{X_jX_k}(\alpha_1-\alpha_q;\nu-\alpha_q) & \cdots & f_{X_jX_k}(0;\nu-\alpha_q) \end{bmatrix} \tag{7}$$

$f_{X_jX_k}$ denoting the jk element of $\mathbf{f}_X$. The relation $\mathbf{f}_{\mathbf{X}}(\alpha;\nu) = \mathbf{A}\mathbf{f}_{\mathbf{S}}(\alpha;\nu)\mathbf{A}^*$ implies that $\mathbf{C}(\nu) = (\mathbf{A}\otimes\mathbf{I}_q)\mathbf{D}(\nu)(\mathbf{A}^*\otimes\mathbf{I}_q)$ where $\mathbf{D}$ is defined similar to $\mathbf{C}$ but with $f_{S_jS_k}$ (the jk element of $\mathbf{f}_S$) in place of $\mathbf{f}_X$, $\mathbf{I}_q$ is the identity matrix of order q and $\otimes$ denotes the Kronecker product:

$$\mathbf{A}\otimes\mathbf{M} = \begin{bmatrix} A_{11}\mathbf{M} & A_{12}\mathbf{M} & \cdots \\ A_{21}\mathbf{M} & A_{22}\mathbf{M} & \cdots \\ \vdots & \vdots & \ddots \end{bmatrix},$$

A_{ij} being the elements of $\mathbf{A}$. The independence of the sources implies that the matrix $\mathbf{D}$ is block diagonal ($\mathbf{D}_{jk} = \mathbf{0}$ except when $j = k$). Thus our idea is to find a separation matrix $\mathbf{B}$ such that $\mathbf{B}\otimes\mathbf{I}_q$ block diagonalizes all the matrices $\mathbf{C}(\nu)$ in the sense that the matrices $(\mathbf{B}\otimes\mathbf{I}_q)\mathbf{C}(\nu)(\mathbf{B}^*\otimes\mathbf{I}_q)$ are block diagonal (of block size q) for all ν.

In practice, the matrices $\mathbf{C}(\nu)$ have to be replaced by their estimators $\hat{\mathbf{C}}(\nu)$. This estimator is naturally built from the estimators $\hat{f}_{\mathbf{X}}(\alpha;\nu)$ of $f_{\mathbf{X}}(\alpha;\nu)$, defined similarly as in (5) with $\hat{R}(\alpha;\tau)$ replaced by $\hat{R}_{\mathbf{X}}(\alpha;\tau)$, the estimator of $R_{\mathbf{X}}(\alpha;\tau)$. The last estimator is defined similarly as in (4) with $X(t)$ replaced by $\mathbf{X}(t)$. As the lag window k_M in (5) has the effect of a smoothing, the (cyclic) spectral density estimator at a given frequency actually does not estimate the spectral density at this frequency but the average density over a frequency band centered around it. Therefore, we shall limit ourselves to the matrices $\hat{\mathbf{C}}(\nu)$ for ν on some finite grid, so that we have only a finite set of matrices to be block diagonalized. The spacing of the grid would be directly related to the resolution of the spectral estimator. Of course, since the $\hat{\mathbf{C}}(\nu)$ are not exactly equal to $\mathbf{C}(\nu)$, one cannot block diagonalize them exactly but only approximately, according to some block diagonality measure, which will be introduced below.

It is important that the estimator $\hat{\mathbf{C}}(\nu)$ be non negative as $\mathbf{C}(\nu)$ is. One can ensure that this is the case regardless of the data, by chosing the (real) window k_M in (5) such that $\sum_\tau k_M(\tau)e^{-2\pi\nu\tau}$ is real and non negative for all ν. Indeed, there then exists a real window $k_M^{1/2}$ (not unique) such that $\sum_\tau k_M(\tau)e^{-2\pi\nu\tau} = |\sum_\tau k_M^{1/2}(\tau)e^{-2\pi\nu\tau}|^2$ or $k_M(\tau) = \sum_u k_M^{1/2}(u-\tau)k_M^{1/2}(u)$. Therefore

$$\hat{f}_{\mathbf{X}}(\alpha;\nu)=\frac{1}{T}\sum_{\tau}\left[\sum_{u}k_M^{1/2}(u-\tau)k_M^{1/2}(u)\right]\left[\sum_{v}\tilde{\mathbf{X}}(v+\tau)\tilde{\mathbf{X}}^*(v)e^{-i2\pi\alpha v}\right]e^{-i2\pi\nu\tau}$$

where $\tilde{\mathbf{X}}(t)=\mathbf{X}(t)$ for $1\leq t\leq T$, $=\mathbf{0}$ otherwise. The last right hand side equals, after summing up with respect to τ: $T^{-1}\sum_u\sum_v(k_M\star\tilde{\mathbf{X}}_\nu)(v+u)\;k_M^{1/2}(u)\mathbf{X}^*(v)e^{i2\pi(\nu-\alpha)v}$ where $\tilde{\mathbf{X}}_\nu(t)=\tilde{\mathbf{X}}(t)e^{-i2\pi\nu t}$ and $\star$ denotes the convolution. Let $t=u+v$ and summing up again first with respect to u, one gets

$$\hat{f}_{\mathbf{X}}(\alpha;\nu)=\frac{1}{T}\sum_{t}(k_M\star\tilde{\mathbf{X}}_\nu)(t)\;(k_M^{1/2}\star\mathbf{X}^*_{\nu-\alpha})(t).$$

This formula shows that $\hat{\mathbf{C}}(\nu)$ is the sample covariance of certain vector sequence, hence is non negative, and can be used for the calculation of $\hat{\mathbf{C}}(\nu)$.

4 Joint Block Approximate Diagonalization

The separation method in previous section leads to the problem of joint approximate block diagonalizing a set of positive definite block matrices $\hat{\mathbf{C}}(\nu_m), m=1,\ldots,M$, of block size q, by a matrix of the form $\mathbf{B}\otimes\mathbf{I}_q$. Following [4] we take as the measure of block diagonality of a Hermitian non negative block matrix $\mathbf{M}$: $(1/2)[\log\det\mathrm{Diag}(\mathbf{M})-\log\det(\mathbf{M})]$ where Diag denotes the operator which builds a bloc diagonal matrix from its argument. This measure is always positive and can be zero if and only if the matrix $\mathbf{M}$ is block diagonal. Indeed, each diagonal block $\mathbf{M}_{ii}$ of $\mathbf{M}$, being non negative, can be diagonalized by a unitary matrix $\mathbf{U}_i$. Thus the matrices $\mathbf{U}_i\mathbf{M}_{ii}\mathbf{U}_i^*$ are diagonal with diagonal elements being also those of $\mathbf{UMU}^*$ where $\mathbf{U}$ is the block diagonal matrix with diagonal block $\mathbf{U}_i$. Hence by the Hadamard inequality [12], $\prod_i\det(\mathbf{U}_i\mathbf{M}_{ii}\mathbf{U}_i^*)\geq\det\mathbf{UMU}^*$ with equality if and only if $\mathbf{UMU}$ is diagonal. This yields the announced result, since the right and left hand sides of the above inequality are no other than $\det\mathrm{Diag}(\mathbf{M})$ and $\det(\mathbf{M})$, and $\mathbf{UMU}^*$ diagonal is the same as $\mathbf{M}$ is block diagonal.

Therefore we consider the joint block diagonality criterion

$$\frac{1}{2}\sum_{m=1}^{M}\{\log\det\mathrm{Diag}[(\mathbf{B}\otimes\mathbf{I}_q)\hat{\mathbf{C}}(\nu_m)(\mathbf{B}\otimes\mathbf{I}_q)]-\log\det[(\mathbf{B}\otimes\mathbf{I}_q)\hat{\mathbf{C}}(\nu_m)(\mathbf{B}^*\otimes\mathbf{I}_q)]\}. \tag{8}$$

Note that the last term in the above curly bracket $\{\ \}$ may be replaced by $2q\log\det|\mathbf{B}|$ since these two terms differ by $\log\det[\mathbf{C}(\nu_m)]$ which does not depend on $\mathbf{B}$.

The algorithm in [13] can be adapted to solve the above problem. For lack of space, we here only describe how it works. Starting from a current value of $\mathbf{B}$, it consists in performing successive transformations, each time on a pair of rows of $\mathbf{B}$, the i-th row $\mathbf{B}_{i\cdot}$ and the j-th row $\mathbf{B}_{j\cdot}$ say, according to

$$\begin{bmatrix}\mathbf{B}_{i\cdot}\\ \mathbf{B}_{j\cdot}\end{bmatrix}\leftarrow\mathbf{T}_{ij}\begin{bmatrix}\mathbf{B}_{i\cdot}\\ \mathbf{B}_{j\cdot}\end{bmatrix},$$

where $\mathbf{T}_{ij}$ is a 2×2 non singular matrix, chosen such that the criterion is decreased and whose expression is given later. Once this is done, the procedure is repeated with another pair of rows. The processing of all the $K(K-1)/2$ is called a sweep. The algorithm consists of repeated sweeps until convergence is achieved. Put

$$g_{ij} = \sum_{m=1}^{M} \frac{1}{Mq} \mathrm{tr}[\mathbf{C}_{ii}^{-1}(m;\mathbf{B})\mathbf{C}_{ij}(m;\mathbf{B})], \qquad 1 \le i \ne j \le K, \tag{9}$$

$$\omega_{ij} = \sum_{m=1}^{M} \frac{1}{Mq} \mathrm{tr}[\mathbf{C}_{ii}^{-1}(m;\mathbf{B})\mathbf{C}_{jj}(m;\mathbf{B})], \qquad 1 \le i \ne j \le K.$$

where $\mathbf{C}_{ij}(m;\mathbf{B})$ stands for the ij block of $(\mathbf{B} \otimes \mathbf{I}_q)\mathbf{C}(\nu_m)(\mathbf{B}^* \otimes \mathbf{I}_q)$ for short. The matrix is $\mathbf{T}_{ij}$ given by

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} - \frac{2}{1 + h_{ij}h_{ji} - h_{ij}^* h_{ji}^* + \sqrt{(1 + h_{ij}h_{ji} - h_{ij}^* h_{ji}^*)^2 - 4h_{ij}h_{ji}}} \begin{bmatrix} 0 & h_{ij} \\ h_{ji} & 0 \end{bmatrix}$$

where h_{ij} and h_{ji} are the solution of

$$\begin{bmatrix} \omega_{ij} & 1 \\ 1 & \omega_{ji} \end{bmatrix} \begin{bmatrix} h_{ij} \\ h_{ji}^* \end{bmatrix} = \begin{bmatrix} g_{ij} \\ g_{ji}^* \end{bmatrix}.$$

Note 1. In the case where the signal $\mathbf{X}(t)$ is real, $\hat{\mathbf{R}}_{\mathbf{X}}(\alpha;\tau) = \hat{\mathbf{R}}_{\mathbf{X}}^T(-\tau)e^{i2\pi\alpha\tau}$, ^{T} denoting the transpose, hence $\mathbf{f}_{\mathbf{X}}(\alpha;-\nu) = \mathbf{f}_{\mathbf{X}}^T(\alpha;\alpha+\nu)$. It follows that

$$f_{X_jX_k}(\alpha_m - \alpha_l; -\nu - \alpha_l) = f_{X_kX_j}(\alpha_m - \alpha_l; \nu + \alpha_m).$$

We already know that if $\mathcal{A}$ contain α it must contain $-\alpha$. Thus it is of interest to choose $\alpha_1 = 0$ and $\alpha_j = -\alpha_{q+2-j}$, $2 \le j \le q$ (which implies that q is odd, unless $1/2 \in \mathcal{A}$, in this case q may be even with $\alpha_{q/2+1} = 1/2$)[1]. Then the above right hand side can be written as $f_{X_kX_j}(\alpha_{q+2-l} - \alpha_{q+2-m}; \nu - \alpha_{q+2-m})$, with $\alpha_{q+1} = 0$ by convention. Therefore by (7): $\mathbf{C}_{jk}(-\nu) = \mathbf{\Pi}\mathbf{C}_{kj}^T(\nu)\mathbf{\Pi}^T$ for some permutation matrix $\mathbf{\Pi}$, hence $\mathbf{C}(-\nu) = (\mathbf{I}_K \otimes \mathbf{\Pi})\mathbf{C}^T(\nu)(\mathbf{I}_K \otimes \mathbf{\Pi}^T)$. It follows that for a *real* matrix $\mathbf{B}$

$$(\mathbf{B} \otimes \mathbf{I}_q)\mathbf{C}(-\nu)(\mathbf{B}^* \otimes \mathbf{I}_q) = (\mathbf{I}_K \otimes \mathbf{\Pi})[(\mathbf{B} \otimes \mathbf{I}_q)\mathbf{C}(\nu)(\mathbf{B}^* \otimes \mathbf{I}_q)]^T(\mathbf{I}_K \otimes \mathbf{\Pi}^T),$$

and thus the measure of block diagonality of the matrix in the above left hand side is the same as that of $(\mathbf{B} \otimes \mathbf{I}_q)\mathbf{C}(\nu)(\mathbf{B}^* \otimes \mathbf{I}_q)$. It is then of interest to consider a grid of frequencies $\nu_1, \ldots, \nu_M$ with M even and $\nu_m = -\nu_{M+1-m} \bmod 1$, so as to reduce the number of matrices to be block diagonalized by half, since the term corresponding to ν_m in (8) can be grouped with the one corresponding to ν_{M+1-m}. One may take $\nu_m = (m - 1/2)/M$ which yield a regular grid of spacing $1/M$.

Note 2. In the case where the signals are real, the matrix $\mathbf{B}$ must be constrained to be real, that is the minimization of (8) must be done over the set of real matrices. It can be shown that the algorithm is the same as before but the g_{ij} are now defined as the real part of the right hand side of (9).

[1] $\{\alpha_1, \ldots, \alpha_q\}$ need not be equal to $\mathcal{A}$ but can be a subset of $\mathcal{A}$.

5 Some Simulation Examples

We consider two cyclostationary sources constructed as Gaussian stationary autoregressive (AR) processes of second order, modulated with sine waves $\cos(\alpha_2 \pi t)$ and $\cos(\alpha_3 \pi t)$ respectively. Thus they have cycle frequencies $0, \pm\alpha_2$ and $0, \pm\alpha_3$ respectively. We take $\alpha_2 = 0.3/\pi = 0.0955$ and $\alpha_3 = 0.7/\pi = 0.2228$ (the same as in [9]). The AR coefficients are $1.9\cos(0.16\pi), -0.95^2$ and $\cos(0.24\pi), -0.5^2$ for the first and second sources, respectively. This corresponds to the AR polynomials with roots $0.95e^{\pm i0.16\pi}$ and $0.5e^{\pm i0.24\pi}$ respectively.

Four hundred series of length 256 are generated for each source. The 2 sources are mixed according to the mixing matrix $\mathbf{A} = \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}$ and our method is applied to obtain the separation matrix $\mathbf{B}$. The number of positive frequency bins is set to 4. To quantify the quality of the separation, we introduce two contamination coefficients c_{12} and c_{21} defined as follows. First the global matrix $\mathbf{G} = \mathbf{BA}$ is formed, then its rows is eventually permuted such that $|G_{11}G_{22}| \geq |G_{12}G_{21}|$, G_{ij} denoting the elements of $\mathbf{G}$. Finally $c_{12} = G_{12}/G_{11}$ and $c_{21} = G_{21}/G_{22}$.

Table 1 shows the mean square of the contamination coefficients and of their products, all multiplied by 256 which is the length of the observed series (since the variance of the estimator should be asymptotically inversely proportional to this length). The mean number of iterations is also listed. For comparison, the values for the stationary method in [4] is also given. This method amounts to running our algorithm with no cycle frequency: $q = 1$ and $\alpha_1 = 0$, which means that one just ignore the cyclostationarity of the sources and considers them as stationary (with spectrum being the average spectrum over time). It can be seen that cyclostationary method yields better results than the stationary method. However, the algorithm converges a little more slowly and each iteration is also more costly computationally.

Table 1. Mean square of the contamination coefficients and of their products and mean number of iterations, obtained from the cyclostationary and stationary methods. The sources are modulated AR processes.

	256(mean c_{12}^2)	256(mean c_{21}^2)	256(mean $c_{12}c_{21}$)	mean # iterations
cyclostationary method	0.3707	0.0310	0.0010	5.86
stationary method	0.5513	0.1250	-0.0628	3.97

In a second test, we consider two cyclostationary sources constructed as (temporally) independent Gaussian processes of unit variance, modulated in the same way as before. Thus the sources are uncorrelated but have variance varying periodically. Therefore, the stationary methods, which amount to considers the sources as stationary with spectrum being the average spectrum over time, would fail since the average sources spectra are constant. Table 2 compares the results of the cyclostationary and stationary methods. It can be seen the stationary method fails completely, as expected. The cyclostationary still works reasonably well, although less well than in the case where the sources are modulated AR processes. The "nonstationarity" method in [6] is also not suitable since

the variance function vary too fast. Indeed, the variance function of the sources have frequencies α_1 and α_2 respectively, which corresponds to the periods $1/\alpha_1 = \pi/0.3 = 10.472$ and $1/\alpha_1 = \pi/0.7 = 4.4880$. Thus in order to "see" the variation of the source variances one has to estimate them in a moving window of size less than 4 which is to short.

Table 2. Mean square of the contamination coefficients and of their products and mean number of iterations, obtained from the cyclostationary and stationary methods. The sources are modulated independent Gaussian processes of unit variance.

	256(mean c_{12}^2)	256(mean c_{21}^2)	256(mean $c_{12}c_{21}$)	mean # iterations
cyclostationary method	0.6638	0.6131	-0.2586	8.27
stationary method	74.4639	72.8648	-72.5062	4.43

References

1. Cardoso, J.F.: Blind signal separation: statistical principles. Proceedings of the IEEE 9, 2009–2025 (1998)
2. Tong, L., Soon, V.C., Huang, Y.F., Liu, R.: Amuse: A new blind identification algorithm. In: Proc. IEEE ISCAS, New Orleans, LA, USA, pp. 1784–1787. IEEE Computer Society Press, Los Alamitos (1990)
3. Belouchrani, A., Meraim, K.A., Cardoso, J.F., Moulines, E.: A blind source separation technique based on second order statistics. IEEE Trans. on Signal Processing 45, 434–444 (1997)
4. Pham, D.T.: Blind separation of instantaneous mixture of sources via the Gaussian mutual information criterion. Signal Processing 81, 850–870 (2001)
5. Matsuoka, K., Ohya, M., Kawamoto, M.: A neural net for blind separation of nonstationary signals. Neural networks 8, 411–419 (1995)
6. Pham, D.T., Cardoso, J.F.: Blind separation of instantaneous mixtures of non stationary sources. IEEE Trans. Signal Processing 49, 1837–1848 (2001)
7. Liang, Y.C., Leyman, A.R., Soong, B.H.: Blind source separation using second-order cyclic-statistics. In: First IEEE Signal Processing Workshop on Advances in Wireless Communications, Paris, France, pp. 57–60. IEEE Computer Society Press, Los Alamitos (1997)
8. Ferreol, A., Chevalier, P.: On the behavior of current second and higher order blind source separation methods for cyclostationary sources. IEEE Trans. on Signal Processing 48, 1712–1725 (2000)
9. Abed-Meraim, K., Xiang, Y., Manton, J.H., Hua, Y.: Blind source-separation using second-order cyclostationary statistics. IEEE Trans. on Signal Processing 49, 694–701 (2001)
10. Ferreol, A., Chevalier, P., Albera, L.: Second-order blind separation of first- and second-order cyclostationary sources-application to AM, FSK, CPFSK, and deterministic sources. IEEE Trans. on Signal Processing 52, 845–861 (2004)
11. Dehay, D., Hurd, H.L.: Representation and estimation for periodically and almost periodically correlated random processes. In: Gardner, W.A. (ed.) Cyclostationarity in Communications and Signal Processing, IEEE Press, Los Alamitos (1993)
12. Cover, T., Thomas, J.: Elements of Information Theory. Wiley, New-York (1991)
13. Pham, D.T.: Joint approximate diagonalization of positive definite matrices. SIAM J. on Matrix Anal. and Appl. 22, 1136–1152 (2001)