Sequential Blind Source Separation Based Exclusively on Second-Order Statistics Developed for a Class of Periodic Signals

Maria G. Jafari, *Member, IEEE*, Wenwu Wang, *Member, IEEE*, Jonathon A. Chambers, *Senior Member, IEEE*, Tetsuya Hoya, *Member, IEEE*, and Andrzej Cichocki, *Senior Member, IEEE*

Abstract—**A sequential algorithm for the blind separation of a class of periodic source signals is introduced in this paper. The algorithm is based only on second-order statistical information and exploits the assumption that the source signals have distinct periods. Separation is performed by sequentially converging to a solution which in effect diagonalizes the output covariance matrix constructed at a lag corresponding to the fundamental period of the source we select, the one with the smallest period. Simulation results for synthetic signals and real electrocardiogram recordings show that the proposed algorithm has the ability to restore statistical independence, and its performance is comparable to that of the equivariant adaptive source separation (EASI) algorithm, a benchmark high-order statistics-based sequential algorithm with similar computational complexity. The proposed algorithm is also shown to mitigate the limitation that the EASI algorithm can separate at most one Gaussian distributed source. Furthermore, the steady-state performance of the proposed algorithm is compared with that of EASI and the block-based second-order blind identification (SOBI) method.**

Index Terms—**Blind source separation, fetal electrocardiogram extraction, periodic signals, second-order statistics.**

I. INTRODUCTION

OVER the past decade blind source separation (BSS) has received much research attention because of its potential applicability to a wide range of problems, spanning disciplines as diverse as communications, geophysical exploration, airport surveillance and medical signal processing [1]–[3]. The problem arises when multidimensional observations, generated when a set of signals are mixed by passage through an unknown medium, must be processed to recover the original sources without the benefit of any *a priori* knowledge about the mixing operation or the sources themselves.

Several approaches have been developed for the solution of the blind source separation problem, which are generally based on the second or higher order statistics of the data. Higher order methods have generally represented the preferred approach to solving BSS problems, since they require few assumptions aside from the statistical independence of the sources, thus addressing the problem in a truly blind context. They are effectively based on both second- and higher order statistics, typically involving a two-step procedure in which whitening precedes rotation, and have resulted in a large number of algorithms (see, for instance, [4]–[9]). Purely second-order statistics methods, on the other hand, have the advantage of requiring shorter data records due to their reduced small sample estimation errors, and do not limit the number of Gaussian sources that can be separated to one [32]–[41]. However, they operate in a semiblind setting, since their derivation usually requires that certain additional assumptions are made on the nature of the original signals, such as statistical nonstationarity of the sources, presence of time correlations in stationary signals, or cyclostationarity; nonetheless, such information is available in, for example, certain biomedical applications, and should be exploited.

Cyclostationary sources have the useful property of spectral redundancy, which arises because of correlation being present between spectral components situated apart from each other, at a distance related to the cyclic frequency. Thus, addressing the BSS problem for cyclostationary sources has the advantage of adding frequency diversity to the spatial diversity already present due to recording the measurements with an array of sensors [13]–[17]. This type of approach does require, of course, knowledge of the cyclic frequencies, or their estimation.

Several block-based methods exist that exploit the temporal correlations of the source signals, and perhaps the best known is the second-order blind identification (SOBI) algorithm [18]. As is usually the case with batch algorithms, the first stage of SOBI entails prewhitening the data; this is generally followed by the joint diagonalization of a set of covariance matrices at different time lags, thus potentially allowing separation of several Gaussian sources. The computational complexity of this algorithm is substantially greater than sequential algorithms due to the need for diagonalization of a number of sample covariance matrices and therefore is not considered further in this work.

Other algorithms that are essentially based on a similar principle can be found in [19] and [20].

In this paper, we introduce a new sequential algorithm for the blind separation of source signals with distinct periods, based only on second-order statistical information. This work is motivated by the observation that the majority of measurements obtained from many biomedical applications exhibit some degree of periodicity. For instance, the human heart contracts at regular intervals, thereby sending the blood through the body; the (normal) electroencephalogram recorded from the scalp is characterized by rhythmic "background activity," and many disorders, such as certain types of epilepsy and Creutzfeld–Jakob disease, are accompanied by patterns that are periodic or almost periodic.

Source separation is performed by sequentially converging to a solution which in effect diagonalizes the output covariance matrices, at a lag τ not equal to zero, but corresponding to the fundamental period of periodicity of the source signal with the smallest period. Besides the computational simplicity, another major advantage of this algorithm is that empirical selection of the lag, or lags, of the covariance matrices to diagonalize is avoided because it is dictated by the fundamental period of one of the sources. This implies that *a priori* knowledge about at least one source is required. Fortunately, in many applications such as in biomedical signal processing, this type of information is often readily available. Alternatively, when a source is known to be periodic, the fundamental period can be estimated using the periodogram of the mixtures, or methods such as the heart instantaneous frequency (HIF) estimation technique proposed in [21].

Moreover, the algorithm can successfully separate the sources even when the selected lag does not correspond to the exact period, as found in simulations (see Section V-E). In addition, the zero lag output covariance matrix is diagonalized when the sources are correctly estimated, so that there is also no need to perform explicit prewhitening. Thus, the decorrelation and rotation operation are in effect carried out concurrently. It is of interest to note that there is no definitive reason why the whitening and rotation process should be conducted in two steps as suggested in most BSS algorithms. In fact, when additive noise is present, "hard prewhitening" can be a major disadvantage [30]. However, in practice, the covariance matrix at the desired lag is not immediately available, but must be estimated, so to avoid delay and reduce computational complexity the operation is implemented sequentially. In this context, a further issue is addressed in this paper: if the sample output spatial covariance matrix at the desired lag is not positive definite, the direction of descent varies, leading to oscillations in the coefficients of the estimated separating matrix. In fact, we believe that this issue has thus far prevented the derivation of a usable sequential counterpart of SOBI. This problem is resolved here by forcing the adjustment term within the sequential algorithm to be positive definite at every iteration.

Since the algorithm is sequential, we first examine why source separation by periodic whitening is possible, while separation based on the zero lag spatial covariance matrix (conventional decorrelation) does not separate the sources. Secondly, it is shown that following convergence, periodic whitening effectively amounts to the simultaneous diagonalization of all the output autocorrelation matrices evaluated at integer multiples of the fundamental period. The remainder of the paper is organized as follows. The problem that we are addressing is formulated in Section II. The identification conditions for the periodic signals are discussed in Section III, which form the theoretical basis of our proposed approach. To implement the proposed idea, a sequential algorithm, together with a systematic approach for lag selection, is developed in Section IV. Then, the performance of our algorithm is substantially evaluated in Section V, where the results are compared with that of the benchmark sequential algorithm, i.e., equivariant adaptive source separation (EASI), and with the block-based SOBI algorithm. Finally, Section VI concludes the paper.

II. PROBLEM FORMULATION

In this paper, it is sought to exploit the temporal periodic nature of the assumed zero mean source signals $\mathbf{s}(k) \in \mathbb{R}^m$, in order to recover them from a set of available observations $\mathbf{x}(k) \in \mathbb{R}^r$, generated by the mixing model

$$\mathbf{x}(k) = \mathbf{A}\mathbf{s}(k) \qquad (1)$$

where $\mathbf{A} \in \mathbb{R}^{r \times m}$ is a fixed but unknown, full column rank, mixing matrix, and k is the discrete-time index. The separating system then estimates the original source signals according to

$$\mathbf{y}(k) = \mathbf{W}(k)\mathbf{x}(k) \qquad (2)$$

where $\mathbf{y}(k) \in \mathbb{R}^m$ represents the recovered sources, and $\mathbf{W}(k) \in \mathbb{R}^{m \times r}$ are the separating matrices. Throughout this paper, for analytical convenience, real quantities will generally be assumed; generalization to complex values is, however, straightforward. We will assume that the number of sources equals the number of measurements, i.e., $m = r$, the exactly determined problem.

Consider the ith original source signal, which is temporally periodic with fundamental period T_i

$$s_i(k) = s_i(k + nT_i) \qquad (3)$$

where $n \in \mathbb{Z}$, and $\mathbb{Z}$ denotes the set of integers. We assume that $s_i(k)$ is obtained by sampling a periodic analogue signal $s_i(t)$, with period T_A, where it is assumed that $T_A \gg T_s$, the sampling period, and $T_s = 1$. Thus, we define $T_i = \lfloor T_A \rfloor$ to be the fundamental period of $s_i(k)$, where $\lfloor \cdot \rfloor$ represents the floor operator. The error, $T_i - T_A$, will be considered in the simulations sections. This definition of the fundamental period will be assumed in the remainder of the paper as will be the assumption for T_s. Thus, the covariance function of the source signal $s_i(k)$ at integer lag τ, defined by

$$\rho_{s_i}(\tau) = \langle s_i(k)s_i(k + \tau) \rangle = \rho_{s_i}(\tau + nT_i) \qquad (4)$$

where $\langle \cdot \rangle$ denotes statistical expectation, is also periodic with period T_i. Then, the covariance matrix of the source signals, at lag τ, is defined as

$$\mathbf{R}_s(\tau) = \langle \mathbf{s}(k)\mathbf{s}^T(k + \tau) \rangle \qquad (5)$$

where $(\cdot)^T$ represents the transpose operator. At a lag $\tau = T_i$, (5) becomes

$$\mathbf{R}_s(T_i) = \langle \mathbf{s}(k)\mathbf{s}^T(k + T_i)\rangle \tag{6}$$

Assuming that the original sources are mutually statistically uncorrelated, the off-diagonal elements of the matrix in (6) satisfy

$$\langle s_j(k)s_i(k + T_i)\rangle = 0, \forall j \neq i \tag{7}$$

while, in general, the value of the diagonal elements will depend on the fundamental period of the sources. Hence, by virtue of (4) and (7), and with the assumption that all the sources are unity variance, the covariance function of the source $s_i(k)$ becomes

$$\langle s_i(k)s_i(k + T_j)\rangle$$
$$= \begin{cases} \langle s_i(k)s_i(k)\rangle = 1, & \text{for } T_j = nT_i, n \in \mathbb{Z} \\ \rho_{s_i}(T_j), & \text{otherwise} \end{cases}. \tag{8}$$

As a result, the source covariance matrix at lag T_i takes the form

$$\mathbf{R}_s(T_i) = \mathbf{\Lambda}_s(T_i) \tag{9}$$

where the elements of $\mathbf{\Lambda}_s(T_i)$ are defined in (8). The matrix $\mathbf{\Lambda}_s(T_i)$ is the eigenvalue matrix of $\mathbf{R}_s(T_i)$, such that its diagonal entries, λ_i, $i = 1, \ldots, m$, represent the eigenvalues of the source covariance matrix at $\tau = T_i$. Since the samples of the covariance function $\rho_{s_i}(T_j)$ can be both positive and negative, as well as zero, the eigenvalues of the source covariance matrix are not guaranteed to ensure that $\mathbf{R}_s(T_i)$ is positive definite. In Theorem 1, for the sake of simplicity, we assume that the source signals are sinusoidal waveforms and that T_i, the fundamental period of source $s_i(k)$, is the smallest period, such that $T_i < T_j, \forall j \in \{1, 2, \ldots, i - 1, i + 1, \ldots, m\}$. Then, due to the periodicity of the covariance function, the above discussion leads to the following.

Theorem 1: Provided that 1) there exist m discrete-time wide-sense stationary sinusoidal signals, i.e., when $T_s = 1, s_i(k) = \sqrt{2}\sin((2\pi k/T_i) + \phi_i)$, with $T_i \in [2, \infty)$ and $\phi_i \in [-\pi, \pi]$, and $s_j(k) = \sqrt{2}\sin((2\pi k/T_j) + \phi_j), \forall j \in \{1, 2, \ldots, i - 1, i + 1, \ldots, m\}$, where T_j are distinct, and $\phi_j \in [-\pi, \pi]$ are statistically independent uniformly distributed random variables and 2) $\forall j, l \in \{1, 2, \ldots, i - 1, i + 1, \ldots, m\}, \cos(2\pi T_i/T_j) \neq \cos(2\pi T_i/T_l) \neq 0$, then the eigenvalues of the source covariance matrix $\mathbf{R}_s(T_i)$ are nonzero and distinct, as follows:

$$\lambda_i \neq \lambda_j \neq 0. \tag{10}$$

Proof: Straightforward, by contradiction. If $\exists j \in \{1, 2, \ldots, i - 1, i + 1, \ldots, m\}$, for which $\cos(2\pi T_i/T_j) = 0$, e.g., when $(T_i/T_j) = (1/4)$, then from (8), $\rho_{s_j}(T_i) = 0$ and hence $\lambda_j = 0$, in violation of the assumption in Theorem 1; or if $\exists j, l \in \{1, 2, \ldots, i - 1, i + 1, \ldots, m\}$, for which $\cos(2\pi T_i/T_j) = \cos(2\pi T_i/T_l)$, then $\lambda_j = \lambda_l$ in violation of the assumption in Theorem 1. ∎

It should be noted that the fact that the eigenvalues are distinct does not imply that source separation will be straightforward. When the eigenvalues are close in value to each other, this may

undermine the robustness of an algorithm. Theorem 1 will now be generalized to the case of the class of periodic discrete-time sources which satisfy Corollary 1.

Corollary 1: The eigenvalues of the source covariance matrix $\mathbf{R}_s(T_i)$ are nonzero and distinct, provided the following:

1) there exist m unity variance, discrete-time source signals with distinct periods;
2) $\rho_{s_j}(T_i) \neq 0, \forall j \in \{1, 2, \ldots, i - 1, i + 1, \ldots, m\}$;
3) $\rho_{s_j}(T_i) \neq \rho_{s_q}(T_i), \forall j, q \in \{1, 2, \ldots, i - 1, i + 1, \ldots, m\}$.

In the remainder of this paper, we will assume that the eigenvalues of the source covariance matrix at the delay $\tau = T_i$ are distinct and nonzero, and that T_i is the smallest period, such that $T_i < T_j, \forall j \in \{1, 2, \ldots, i - 1, i + 1, \ldots, m\}$. Note that other nonperiodic discrete-time signals can have the required eigenstructure in the covariance matrix, but these are not the focus of this paper. Moreover, this result implies that the resultant algorithm will have applicability beyond the separation of purely sinusoidal signals which underpins its practical utility, for example, in biomedical applications.

III. IDENTIFICATION PRINCIPLE

Conventional decorrelation (or whitening) is typically performed by considering the spatial output covariance matrix when $\tau = 0$. In this case, a whitening matrix, $\mathbf{V}(k)$, is sought, such that the output vector is given by

$$\mathbf{z}(k) = \mathbf{V}(k)\mathbf{x}(k) \tag{11}$$

and after convergence of the sequential algorithm, i.e., $\lim_{k \to \infty} \langle \mathbf{V}(k)\rangle = \mathbf{V}^o$, its covariance matrix satisfies

$$\mathbf{R}_z(0) = \lim_{k \to \infty} \langle \mathbf{z}(k)\mathbf{z}^T(k)\rangle = \mathbf{I} \tag{12}$$

and therefore

$$\mathbf{R}_z(0) = \langle \mathbf{V}^o\mathbf{A}\mathbf{s}(k)\mathbf{s}^T(k)\mathbf{A}^T\mathbf{V}^{oT}\rangle \tag{13}$$

which in terms of $\mathbf{U} = \mathbf{V}^o\mathbf{A}$ is

$$\mathbf{R}_z(0) = \mathbf{U}\mathbf{R}_s(0)\mathbf{U}^T. \tag{14}$$

Since

$$\mathbf{R}_s(0) = \langle \mathbf{s}(k)\mathbf{s}^T(k)\rangle = \mathbf{I} \tag{15}$$

and from (12), we have

$$\mathbf{R}_z(0) = \mathbf{U}\mathbf{U}^T = \mathbf{I} \tag{16}$$

which implies that $\mathbf{U}$ is an orthogonal (unitary for complex case) matrix. Effectively, the relations in (12) and (16) indicate that $\mathbf{U}$ cannot achieve separation because of the lack of distinct eigenvalues.

Conversely, when decorrelation is carried out at $\tau = T_i$, an instantaneous output covariance matrix can be formed from

$$\mathbf{R}_y(k, T_i) = \mathbf{P}(k)\mathbf{\Lambda}_s(T_i)\mathbf{P}^T(k + T_i) \tag{17}$$

where the result in (9) has been used, and $\mathbf{P}(k) = \mathbf{W}(k)\mathbf{A}$. Let $\mathbf{R}_y^o(T_i)$ denote the output covariance matrix at convergence of a sequential algorithm, that is

$$\lim_{k \to \infty} \langle \mathbf{R}_y(k, T_i)\rangle = \mathbf{R}_y^o(T_i). \tag{18}$$

By definition, the whitening operation implies that

$$\mathbf{R}_y^o(T_i) = \mathbf{P}^o \mathbf{\Lambda}_s(T_i) \mathbf{P}^{oT} = \hat{\mathbf{D}}(T_i) \qquad (19)$$

where $\hat{\mathbf{D}}(T_i)$ is a diagonal matrix and, since the system (19) is in steady-state, $\mathbf{P}^o = \lim_{k \to \infty} \langle \mathbf{P}(k) \rangle$, (19) implies that $\mathbf{P}^o$ is an orthogonal (unitary for complex case) matrix which can now achieve separation provided that its eigenvalues are distinct and nonzero. Moreover, the columns of $\mathbf{P}^o$ are the eigenvectors of $\mathbf{R}_y^o(T_i)$.[1] This leads to the following theorem.

Theorem 2: Provided that Corollary 1 holds, the eigenvalues of $\mathbf{R}_y^o(T_i)$ are distinct and nonzero.

Proof: Pre- and postmultiplying (19) by $\mathbf{P}^{oT}$ and $\mathbf{P}^o$ respectively, we have

$$\mathbf{P}^{oT} \mathbf{R}_y^o(T_i) \mathbf{P}^o = \mathbf{P}^{oT} \mathbf{P} \mathbf{\Lambda}_s(T_i) \mathbf{P}^{oT} \mathbf{P}^o = \mathbf{\Lambda}_s(T_i) \qquad (20)$$

where $\mathbf{\Lambda}_s(T_i)$ is the diagonal matrix of eigenvalues of $\mathbf{R}_s(T_i)$ (see Section II). Clearly, from (20), $\mathbf{\Lambda}_s(T_i)$ also contains the eigenvalues of $\mathbf{R}_y^o(T_i)$, and therefore, provided that Theorem 1 holds, its diagonal entries are distinct and nonzero. ∎

Next, we show that whitening the output covariance matrix at a lag $\tau = T_i$, simultaneously whitens all the covariance matrices at $\tau = nT_i, n \in \mathbb{Z}$, thus implying that a separating matrix $\mathbf{W}^o$ can be identified, such that the columns of $\mathbf{P}^o = \mathbf{W}^o \mathbf{A}$ are the eigenvectors shared by all output covariance matrices, including $\mathbf{R}_y^o(0)$ because, since the latter converges to the identity matrix, all invertible matrices will diagonalize it.

Theorem 3: Diagonalization of $\mathbf{R}_y^o(T_i)$ effectively amounts to simultaneously diagonalising the covariance matrix at every delay $\tau = nT_i, n \in \mathbb{Z}$.

Its proof is given in the Appendix. This theorem indicates that there exists a separating matrix $\mathbf{W}^o$ such that $\mathbf{P}^o = \mathbf{W}^o \mathbf{A}$ diagonalizes all the matrices $\mathbf{R}_y^o(nT_i), n \in \mathbb{Z}$, provided that its eigenvalues are distinct. This diagonalization corresponds to the simultaneous eigenvalue decomposition (EVD) of $\mathbf{R}_y^o(nT_i)$, and it results in the blind separation of periodic sources, which forms the basis of a sequential algorithm developed in the next section.

IV. SEPARATION ALGORITHM

A sequential algorithm for the separation of periodic signals can be derived by minimizing the following information theoretic criterion which is an extension of the form found in [25, cf. pp. 134–136, 255–258] (see also [10])

$$J(\mathbf{W}(k)) = -\frac{1}{2} \log(\det(\mathbf{W}(k)\mathbf{W}^H(k)))$$
$$+ \frac{1}{2} \sum_{i=1}^{m} \langle |y_i(k) y_i^*(k+\tau)| \rangle \qquad (21)$$

where $|\cdot|$ denotes the absolute value, $(\cdot)^H$ Hermitian transpose, and $(\cdot)^*$ complex conjugate. Further justification of this cost function is given after the derivation of the update (29). Note that we perform the derivation for the complex case to obtain the most general form of the algorithm. Differentiation of the

[1]The matrix $\mathbf{R}_y(k, T_i)$ converges to $\hat{\mathbf{D}}(T_i)$ rather than $\mathbf{\Lambda}_s(T_i)$ because of the indeterminacies of scaling and permutation, which are due to the nonuniqueness of the eigenvectors contained in $\mathbf{P}^o$.

first term on the right-hand side (RHS) of (21) with respect to the separating matrix $\mathbf{W}(k)$ leads to

$$\frac{\partial \log(\det(\mathbf{W}(k)\mathbf{W}^H(k)))}{\partial \mathbf{W}(k)} = 2(\mathbf{W}(k)\mathbf{W}^H(k))^{-1}\mathbf{W}(k) \qquad (22)$$

The second term on the RHS of the cost function is differentiated following the approach in [25, cf. 135]

$$\frac{\partial \langle |y_i(k) y_i^*(k+\tau)| \rangle}{\partial w_{ij}(k)}$$
$$= \left\langle \frac{\partial |y_i(k)|}{\partial y_i(k)} y_i^*(k+\tau) \right\rangle \frac{\partial y_i(k)}{\partial w_{ij}(k)}$$
$$+ \left\langle \frac{\partial |y_i^*(k+\tau)|}{\partial y_i^*(k+\tau)} y_i(k) \right\rangle \frac{\partial y_i^*(k+\tau)}{\partial w_{ij}(k)}$$
$$\approx \langle y_i^*(k+\tau) x_j(k) \rangle + \langle y_i(k) x_j^*(k+\tau) \rangle,$$
$$\forall i, j = \{1, 2, \dots, m\} \qquad (23)$$

where it has been assumed that

$$w_{ij}(k) \approx w_{ij}(k+\tau). \qquad (24)$$

In matrix form, (23) becomes

$$\frac{\partial \sum_{i=1}^{m} \langle |y_i(k) y_i^*(k+\tau)| \rangle}{\partial \mathbf{W}(k)}$$
$$\approx \langle \mathbf{y}(k+\tau)\mathbf{x}^H(k) + \mathbf{y}(k)\mathbf{x}^H(k+\tau) \rangle. \qquad (25)$$

Employing the natural gradient descent method, the gradient of (21) is obtained from (22) and (25) as

$$\nabla J(\mathbf{W}(k))$$
$$= \frac{\partial J(\mathbf{W}(k))}{\partial \mathbf{W}(k)} \mathbf{W}^H(k)\mathbf{W}(k) \qquad (26)$$
$$= -\frac{1}{2}[2(\mathbf{W}(k)\mathbf{W}^H(k))^{-1}(\mathbf{W}(k)\mathbf{W}^H(k))\mathbf{W}(k)$$
$$- \langle \mathbf{y}(k+\tau)\mathbf{x}^H(k) + \mathbf{y}(k)\mathbf{x}^H(k+\tau) \rangle \mathbf{W}^H(k)\mathbf{W}(k)]$$
$$= -\left[\mathbf{I} - \frac{1}{2}\langle \mathbf{y}(k+\tau)\mathbf{x}^H(k)\mathbf{W}^H(k) \right.$$
$$\left. + \mathbf{y}(k)\mathbf{x}^H(k+\tau)\mathbf{W}^H(k) \rangle \right] \mathbf{W}(k)$$
$$= -\left[\mathbf{I} - \frac{1}{2}\langle \mathbf{y}(k+\tau)\mathbf{y}^H(k) + \mathbf{y}(k)\mathbf{y}^H(k+\tau) \rangle \right] \mathbf{W}(k) \qquad (27)$$

where the assumption (24) has again been used, but for convenience the approximation in (25) is not explicitly shown. In terms of the output covariance matrix $\mathbf{R}_y(k, T_i) = \langle \mathbf{y}(k)\mathbf{y}^H(k+T_i) \rangle$, (27) becomes

$$\nabla J(\mathbf{W}(k)) = -\left[\mathbf{I} - \frac{1}{2}\left(\mathbf{R}_y^H(k, T_i) + \mathbf{R}_y(k, T_i)\right)\right] \mathbf{W}(k). \qquad (28)$$

This leads to the following learning rule:

$$\mathbf{W}(k+1) = \mathbf{W}(k) + \eta \left[\mathbf{I} - \frac{1}{2}\left(\mathbf{R}_y^H(k, T_i)\mathbf{R}_y(k, T_i)\right)\right] \mathbf{W}(k) \qquad (29)$$

where η is a positive step-size parameter.

The role of $-(1/2) \log(\det(\mathbf{W}(k)\mathbf{W}^H(k)))$ in the right-hand side (RHS) of (21) is to avoid a zero solution

for $\mathbf{W}(k)$ and this is clear in (29) by the appearance of the identity matrix, which ensures that in steady state, i.e., $\lim_{k\to\infty} E\{\mathbf{W}(k+1)\} = \lim_{k\to\infty} E\{\mathbf{W}(k)\}$, the outputs $y_i(k), i = 1, \ldots, m$, will have unity power. The second term in the RHS of (21) is present to ensure that at the chosen lag τ, the output signals have minimum sum absolute cross correlation values. Its role is better understood from the update (29), which demonstrates in steady state the average (Hermitian) output covariance matrix at lag T_i will be diagonalized.

To analyze the stability, we first show that the cost function (21) is a nonnegative function. Denote by $w_{ij}(k)$ the entries of $\mathbf{W}(k)$, $\lambda_i(k)$ the eigenvalues of $\mathbf{W}(k)\mathbf{W}^H(k)$, according to Hadamard's inequality (see, e.g., [25, cf. p. 255]), we have

$$
\begin{aligned}
\prod_{i=1}^{m} \lambda_i(k) &= \det(\mathbf{W}(k)\mathbf{W}^H(k)) \\
&\le \prod_{i=1}^{m} \sum_{j=1}^{r} |w_{ij}(k)|^2 \\
&\le \prod_{i=1}^{m} \langle |y_i^2(k)| \rangle \\
&\approx \prod_{i=1}^{m} \langle |y_i(k)y_i(k+\tau)| \rangle.
\end{aligned}
$$
(30)

The approximation in the last term of (30) is on the basis of appropriate choice of τ. According to the monotonic property of exponential (logarithmic) functions, it is thereby straightforward to check the nonnegativity of (21). Since update (29) is derived by following the natural gradient learning rule (26), its stability can be readily attained (see, e.g., [25, cf. pp. 168–170]). To this end, we consider its corresponding continuous system (denoted by time index t). Denote $\psi_{ij}, i, j = 1, \ldots, m$, the elements of $(\partial J/\partial \mathbf{W})\mathbf{W}^H$, we obtain from (26) that

$$
\begin{aligned}
\frac{d}{dt} J(\mathbf{y}(t), \mathbf{W}(t)) &= \sum_{i,j} \frac{\partial J}{\partial w_{ij}} \frac{dw_{ij}}{dt} \\
&= -\sum_{i,j} \frac{\partial J}{\partial w_{ij}} \sum_{p} \psi_{ip} w_{pj} \\
&= -\sum_{i,p} \psi_{ip}^2 \le 0
\end{aligned}
$$
(31)

where zero is obtained if and only if $d\mathbf{W}(t)/dt = 0$, which means the solution to $\mathbf{W}$ is an equilibrium of (21). In other words, the learning (29) will force the algorithm to converge to a stationary point. As is convention to sequential algorithms, since our algorithm has the simple form of a recursive update, the stationary points will correspond to the points for which the average adjustment term (of the update equation) goes to zero. That is, the update (29) ceases if $E[\mathbf{I}-(1/2)(\mathbf{R}_y^H(k, T_i)+\mathbf{R}_y(k, T_i))] = 0$. Certainly, the algorithm requires an appropriately small adaptation gain η to converge to a stationary point of the cost function. The same analysis is also applicable to the subsequent modified update (37).

The practical implementation of the algorithm in (29) has a drawback: it may fail to converge when $\mathbf{R}_y(k, T_i)$ is not positive definite, because the nature of the optimization problem may vary during the adaptive procedure, i.e., switching between searching for a minimizer or maximizer within parameter space, causing some of the elements of the estimated separating matrix to oscillate between positive and negative values. A solution to this problem is to replace $\mathbf{R}_y(k, T_i)$ in (29) with $\mathbf{R}_y(k, T_i)\mathbf{S}(\mathbf{y}(k, T_i))$, where (see [25, cf. pp. 321–325] and [31])

$$
[\mathbf{S}(\mathbf{y}(k, T_i))]_{i,j} = \begin{cases} \frac{\langle y_i(k)y_i(k+T_i)\rangle}{|\langle y_i(k)y_i(k+T_i)\rangle|}, & \text{if } i = j \\ 0, & \text{otherwise} \end{cases}.
$$
(32)

Thus, (29) becomes

$$
\begin{aligned}
\mathbf{W}(k+1) &= \mathbf{W}(k) + \eta \Big\{ \mathbf{I} - \frac{1}{2}\Big[\mathbf{R}_y^H(k, T_i)\mathbf{S}(\mathbf{y}(k, T_i)) \\
&\quad + \mathbf{R}_y(k, T_i)\mathbf{S}(\mathbf{y}(k, T_i)) \Big] \Big\} \mathbf{W}(k).
\end{aligned}
$$
(33)

Theorem 4: The matrix $\mathbf{R}_y(k, T_i)\mathbf{S}(\mathbf{y}(k, T_i))$ used in (33) is guaranteed to be positive definite.

We leave the proof to the Appendix. It should be noted moreover that the matrix $\mathbf{R}_y^H(k, T_i)+\mathbf{R}_y(k, T_i)$ is Hermitian, which implies that it does not offer enough degrees of freedom for source separation. On the other hand, the matrix

$$
\begin{aligned}
\mathbf{B}_y(k, T_i) &\triangleq \mathbf{R}_y^H(k, T_i)\mathbf{S}(\mathbf{y}(k, T_i)) \\
&\quad + \mathbf{R}_y(k, T_i)\mathbf{S}(\mathbf{y}(k, T_i))
\end{aligned}
$$
(34)

may not be Hermitian when the elements of $\mathbf{S}(\mathbf{y}(k, T_i))$ have different signs, because it involves the multiplication of both $\mathbf{R}_y(k, T_i)$ and $\mathbf{R}_y^H(k, T_i)$ by $\mathbf{S}(\mathbf{y}(k, T_i))$, so that a sufficient number of degrees of freedom can be preserved for the solution of the BSS problem.

For the practical implementation of (33), the matrix $\mathbf{R}_y(k, T_i)$ is replaced by its instantaneous estimate, given by

$$
\hat{\mathbf{R}}_y(k, T_i) = \mathbf{y}(k)\mathbf{y}^H(k+T_i)
$$
(35)

where $\mathbf{y}(k)$ takes the form of (2) and $\mathbf{y}(k+T_i) = \mathbf{W}(k)\mathbf{x}(k+T_i)$. Likewise, $\mathbf{S}(\mathbf{y}(k, T_i))$ is given by

$$
[\hat{\mathbf{S}}(\mathbf{y}(k, T_i))]_{i,j} = \begin{cases} \frac{y_i(k)y_i(k+T_i)}{|y_i(k)y_i(k+T_i)|}, & \text{if } i = j \\ 0, & \text{otherwise} \end{cases}.
$$
(36)

Therefore, (33) becomes

$$
\begin{aligned}
\mathbf{W}(k+1) &= \mathbf{W}(k) + \eta \Big\{ \mathbf{I} - \frac{1}{2}\Big[\hat{\mathbf{R}}_y^H(k, T_i)\hat{\mathbf{S}}(\mathbf{y}(k, T_i)) \\
&\quad + \hat{\mathbf{R}}_y(k, T_i)\hat{\mathbf{S}}(\mathbf{y}(k, T_i)) \Big] \Big\} \mathbf{W}(k).
\end{aligned}
$$
(37)

And provided the dynamic behavior of $\mathbf{y}(k, T_i)$ is such that $\hat{\mathbf{B}}_y(k, T_i)$ generally retains a nonsymmetric but positive definite form during adaptation, the stationary point of (37), i.e.,

$\lim_{k\to\infty} E\{\mathbf{W}(k+1)\} = \lim_{k\to\infty} E\{\mathbf{W}(k)\}$, leads to a separating solution as shall be shown by simulation.

The constraints being imposed on the magnitude of the recovered signals in (37) are typically useful in resolving the scaling ambiguity associated with BSS,[2] and while they lead to good performance when the sources have fixed magnitudes, they result in numerical instability of the algorithm when the amplitude of the sources varies rapidly, as is often the case with applications such as biomedical measurements or audio signals, where periods of large and small magnitude alternate (see [25, cf. pp. 321–325] and [31]). In such cases, numerical instability can be quite serious, and results in failure of the algorithm. Following the approach in [26], this drawback can be resolved by replacing (37) with the learning rule

$$\mathbf{W}(k+1) = \mathbf{W}(k) + \eta \left\{ \mathbf{\Gamma}(k) - \frac{1}{2} \left[\hat{\mathbf{R}}_y^H(k, T_i)\hat{\mathbf{S}}(\mathbf{y}(k, T_i)) \right. \right.$$
$$\left. \left. + \hat{\mathbf{R}}_y(k, T_i)\hat{\mathbf{S}}(\mathbf{y}(k, T_i)) \right] \right\} \mathbf{W}(k)$$
$$(38)$$

where $\mathbf{\Gamma}(k)$ is selected according to (39), shown at the bottom of the page, so that its role is to eliminate the diagonal elements within the brackets of the RHS term of (38).

An important point to note is that the algorithms proposed here have the ability to separate source signals for any arbitrary value of the lag τ. Thus, although the methods have been derived for the case of temporally periodic sources, provided that the source covariance function is nonzero ($\rho(\tau) \neq 0$, hence avoiding zero eigenvalues), and that the sources have different spectral characteristics (resulting in distinct eigenvalues), the assumption that the lag equals the fundamental period of periodicity can be relaxed.

V. SIMULATION RESULTS

Computer simulations were carried out to illustrate the performance of the proposed method, which was compared to that of the EASI algorithm as a benchmark higher order statistics-based sequential algorithm, given by [8]

$$\mathbf{W}(k+1) = \mathbf{W}(k) + \mu\{\mathbf{I} - \mathbf{y}(k)\mathbf{y}^T(k)$$
$$+ \mathbf{y}(k)\mathbf{f}^T(\mathbf{y}(k)) - \mathbf{f}(\mathbf{y}(k))\mathbf{y}^T(k)\}\mathbf{W}(k) \quad (40)$$

where μ is a positive step-size parameter, and $\mathbf{f}(\mathbf{y}(k))$ is an odd nonlinear function of the output $\mathbf{y}(k)$. Note that, we typically use $[\mathbf{y}(k)]^3$ for sub-Gaussians, $\tanh(\mathbf{y}(k))$ for super-Gaussians. In our experiments, we address all sub-Gaussian signals

---

[2]The ambiguity of scaling is only partly resolved when the normalization condition is imposed on the independent components, since this does not eliminate the sign ambiguity, implying that it is still possible to multiply a source by -1 without altering the mixing model.
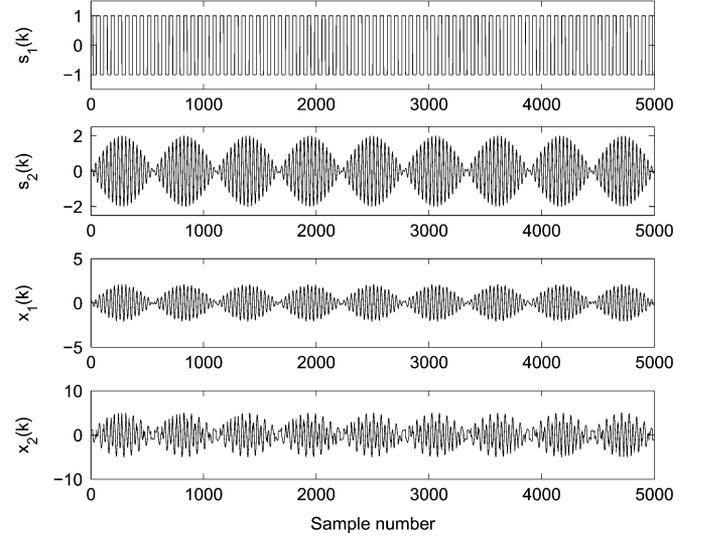

Fig. 1. Original source signals ($s_1(k)$ and $s_2(k)$) and mixed signals ($x_1(k)$ and $x_2(k)$).

except in the ECG simulations. The EASI algorithm explicitly exploits the second- and higher order statistics of the source signals, since the first two terms within the brackets on the RHS of (40) effectively perform prewhitening on the output signals, while the others measure the degree of independence between them.

A. Restoration of Independence

Fig. 1 shows the original sources, and the mixtures obtained with a randomly generated stationary mixing matrix. The sources were a sign, and an amplitude-modulated signal, given by

$$s_1(k) = \text{sgn}(\cos(2\pi f_1 t_s k))$$
$$s_2(k) = \sin(2\pi f_2 t_s k)\sin(2\pi f_3 t_s k)$$

where $t_s = 1 \times 10^{-4}$, $f_1 = 0.1395$, $f_2 = 0.0081$, and $f_3 = 0.27$. Note here, t_s is the sampling period, and f_i ($i = 1, \ldots, 3$) are normalized frequencies.

Separation was performed with (37), when $\eta = 0.003$, $T_i = (f_2 + f_3)^{-1}$ where $(f_2 + f_3)$ is the upper sideband of source $s_2(k)$, and with the EASI algorithm (40), when $\mu = 0.003$. In addition, the mixtures were decorrelated. In Fig. 2, the decorrelated mixtures are denoted by $y_1^D(k)$ and $y_2^D(k)$, while $y_1^E(k)$ and $y_2^E(k)$ represent the signals recovered with the EASI algorithm, and $y_1^P(k)$ and $y_2^P(k)$ are the outputs of the algorithm (37). The results illustrate that, as expected, zero lag decorrelation alone does not separate the sources, whereas the proposed method recovers the original components, and preserves waveform similarity in a similar fashion as the EASI algorithm.

$$[\mathbf{\Gamma}(k)]_{i,j} = \begin{cases} \frac{1}{2} \left[\hat{\mathbf{R}}_y^H(k, T_i)\hat{\mathbf{S}}(\mathbf{y}(k, T_i)) + \hat{\mathbf{R}}_y(k, T_i)\hat{\mathbf{S}}(\mathbf{y}(k, T_i)) \right], & \text{if } i = j \\ 0, & \text{otherwise} \end{cases} \quad (39)$$
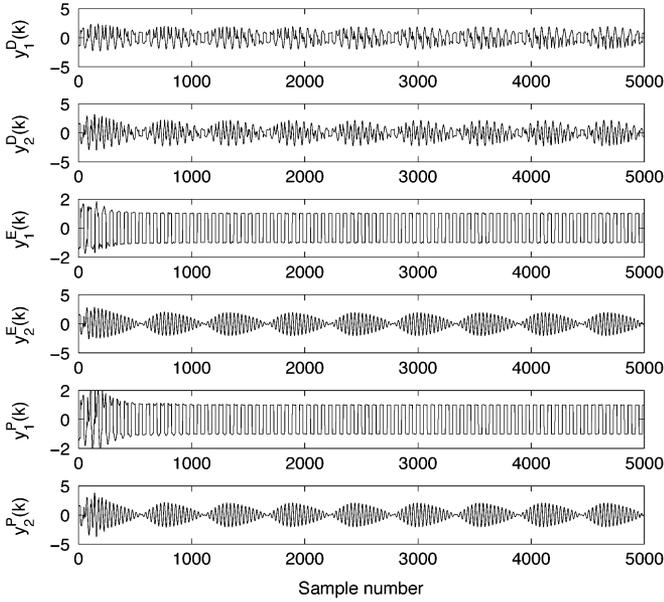
Fig. 2. Decorrelated mixtures ($y_1^D(k)$ and $y_2^D(k)$), and signals recovered with the EASI algorithm ($y_1^E(k)$ and $y_2^E(k)$), and with the algorithm (37) ($y_1^P(k)$ and $y_2^P(k)$).
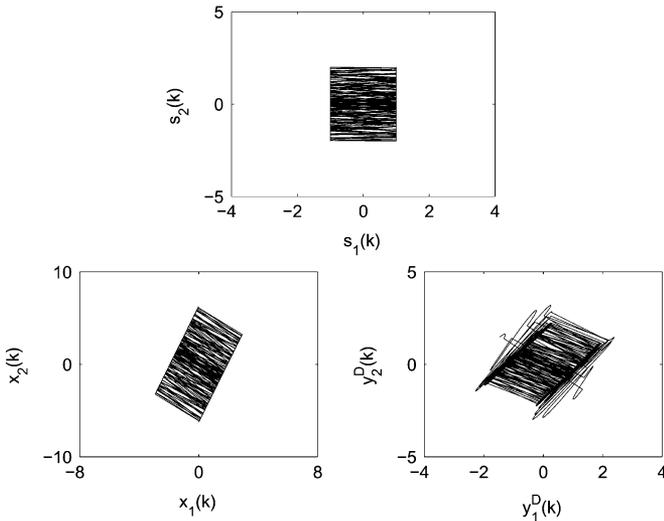

Fig. 3. Scatter plots of the source signals (upper plot), the mixtures (lower left plot), and the mixtures following decorrelation (lower right plot).

The upper plot in Fig. 3 depicts the support of the joint probability density function (pdf) of the source signals, while the support of the joint pdfs of the mixtures and the decorrelated mixtures are shown, respectively, in the lower left and lower right plots. The shape of the joint distribution of the source signals clearly indicates that they are statistically independent, since the knowledge of the value of one does not convey any information about the value of the other. The signals generated by the mixing procedure, x_1 and x_2, are no longer independent, and the joint pdf now has the form of a parallelogram; also, the lower-right plot shows that following zero lag decorrelation the signals are still dependent. The performance of the periodic BSS algorithm
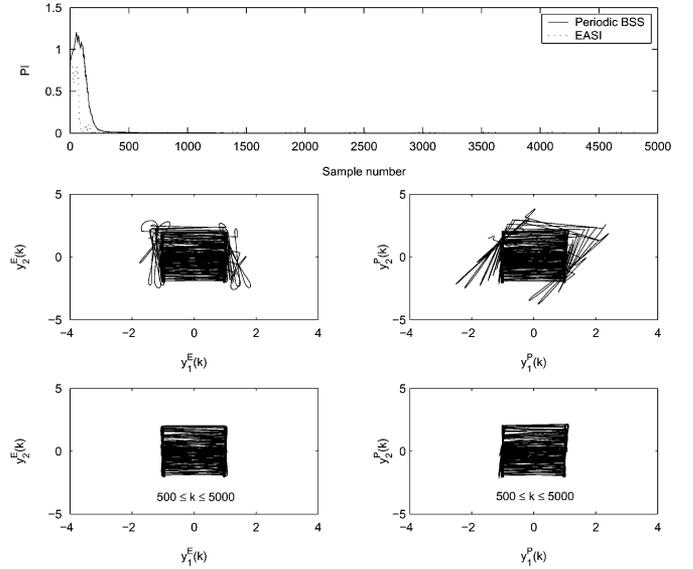

Fig. 4. Performance indexes obtained with the algorithm (37) and with the EASI algorithm (40) (uppermost plot), and scatter plots of the output signals. The middle plots show the scatter plots of the sources separated with the proposed periodic BSS algorithm (right) and with EASI (left) for all the samples, while the lowermost plots represent the scatter plots of the sources recovered with the two algorithms for samples 500 to 5000. Note that only a single trial is conducted in this simulation.

(37) and EASI (40) was compared in terms of the performance index (PI), defined as [27]

$$\mathbf{PI}(k) = \frac{1}{n} \sum_{i=1}^{n} \left\{ \sum_{j=1}^{n} \frac{|p_{ij}|^2}{\max_q |p_{iq}|^2 - 1} \right\}$$
$$+ \frac{1}{n} \sum_{j=1}^{n} \left\{ \sum_{i=1}^{n} \frac{|p_{ij}|^2}{\max_q |p_{qj}|^2 - 1} \right\} \quad (41)$$

where $\mathbf{P}(k) = [p_{ij}] = \mathbf{W}(k)\mathbf{A}$, and n is the number of source signals. Generally, a low PI indicates better performance. The performance indexes for the two methods are compared in the upper plot of Fig. 4, which illustrates that the proposed algorithm is capable of separating the source signals, with a performance comparable to the behavior of EASI. The middle and lower plots in Fig. 4 show the support of the joint probability density function of the output signals. The middle plots represent the support of the joint pdf of the sources separated with the proposed periodic BSS algorithm (right) and with EASI (left) for all the samples, while the lowermost plots illustrate the support of the joint pdf of the outputs for $500 \leq k \leq 5000$. The results show that, unlike conventional pre-whitening, the periodic decorrelation algorithm proposed here has the necessary degrees of freedom to allow it to rotate the joint pdf of the recovered sources, as well as decorrelate the signals, leading to the restoration of statistical independence.

B. Several Gaussian Sources

In this section we consider the performance of the proposed algorithm when more than one source is Gaussian, and compare its behavior to that of the EASI algorithm. Three Gaussian
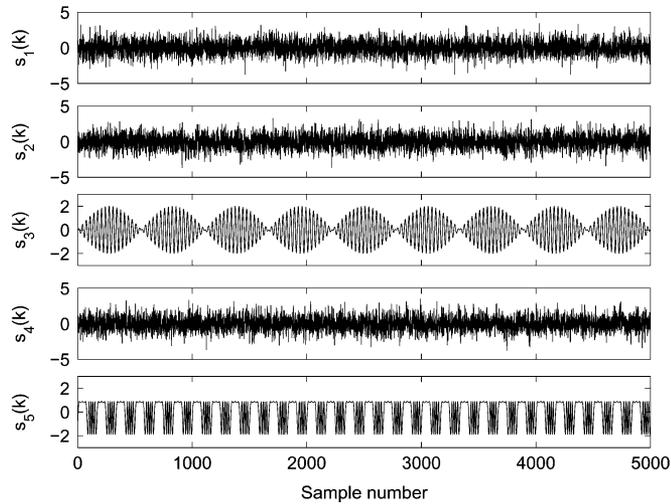
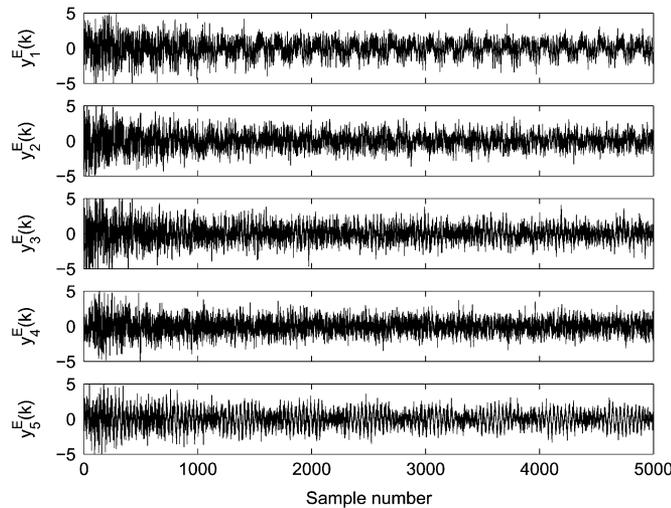Fig. 5. Five original source signals: three Gaussian and two sub-Gaussian.

Fig. 7. Signals recovered with the periodic BSS algorithm (37) when the original sources are as shown in Fig. 5.

Fig. 6. Output signals obtained with the EASI algorithm (40) when the original sources are as shown in Fig. 5.

Fig. 8. Performance index obtained with the proposed periodic BSS algorithm and EASI, averaged over 100 independent trials. The average PIs over the last 3000 samples are 5.6×10^{-3} and 3.9×10^{-3} for the periodic BSS and EASI algorithms respectively.

sources with zero mean and unit variance, and two sub-Gaussian signals (Fig. 5), given by

$$s_1(k) = \sin(2\pi f_2 t_s k) \sin(2\pi f_3 t_s k)$$
$$s_2(k) = \sin(2\pi f_3 t_s k + 6\cos(2\pi f_4 t_s k))$$

where $t_s = 1 \times 10^{-4}$, $f_2 = 0.0081$, $f_3 = 0.27$, and $f_4 = 0.054$, were mixed by a randomly generated 5×5 mixing matrix, and separation was performed with the periodic BSS algorithm (37) with $\eta = 0.003$, $T_i = (f_2 + f_3)^{-1}$, and EASI with $\mu = 10^{-5}$. It was observed that, as expected, the EASI algorithm could not recover the original sources due to the presence of several Gaussian components (Fig. 6), and in particular it was found to fail when the step-size parameter μ was increased. Conversely, the proposed algorithm extracted the two periodic sources, as shown in Fig. 7.

C. Monte Carlo Simulation and Real Signals

The two sources in Fig. 1 and a signal uniformly distributed in $[-1, 1]$ were mixed by a 3×3 mixing matrix, and zero-mean white Gaussian noise was added such that the signal-to-noise
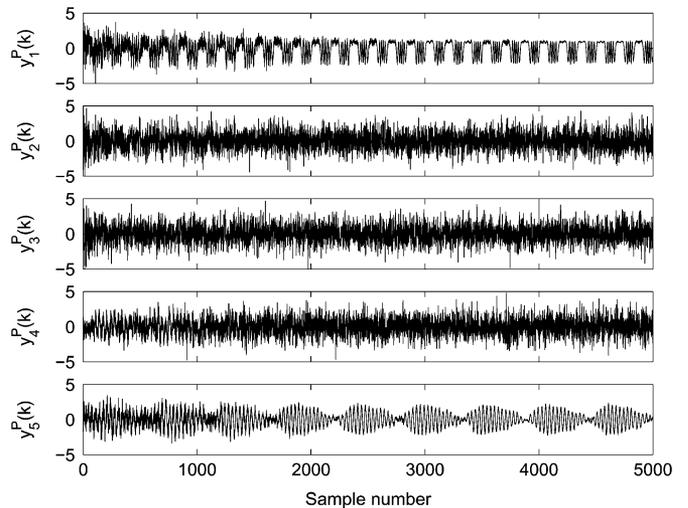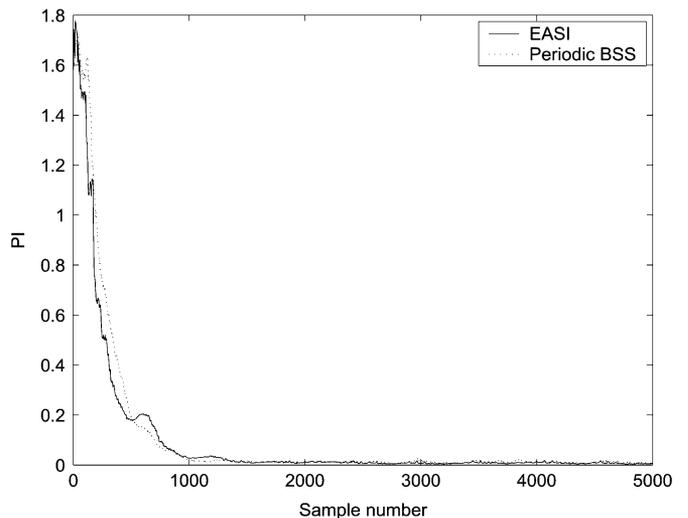
TABLE I
FINAL SEPARATION PERFORMANCE OF THE PROPOSED ALGORITHM AS COMPARED WITH EASI AND SOBI ALGORITHMS

Algorithm	Periodic BSS	EASI	SOBI
Test 1 (PI)	3.08×10^{-5}	2.09×10^{-5}	5.12×10^{-8}
Test 2 (PI)	1.84×10^{-3}	1.80×10^{-3}	1.23×10^{-3}

ratio was 5 dB. Fig. 8 illustrates the performance index resulting when separation was performed with the proposed periodic BSS algorithm (37), and the EASI algorithm (40), averaged over 100 independent trials. It shows that the behavior of the proposed method is similar to that of EASI, both during and following initial convergence.

To compare the final separation performance, we conducted two groups of tests. In the first test, the two sources in Fig. 1 were mixed together using a randomly selected 2×2 mixing matrix.
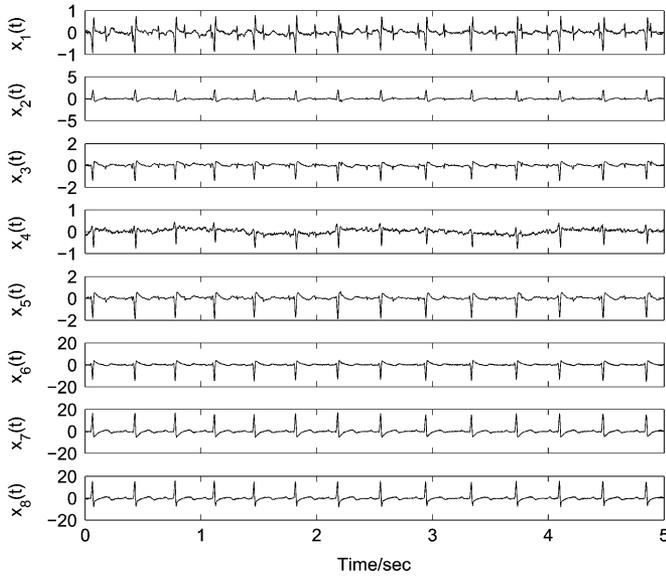
Fig. 9. Eight-channel cutaneous potential recording from a pregnant woman. The signals denoted $x_1(t)$–$x_5(t)$ were recorded from the abdominal area, while the lowermost recordings $x_6(t)$–$x_8(t)$ were obtained from the thoracic area.

Fig. 10. Signals recovered with the algorithm (38), showing fetal components ($y_1^P(t)$ and $y_3^P(t)$), maternal ECG components ($y_6^P(t)$, $y_7^P(t)$, and $y_8^P(t)$), extracted respiratory motion ($y_4^P(t)$) and noise sources ($y_2^P(t)$ and $y_5^P(t)$).

In the second test, besides these two sources, we chose another source signal as used in Fig. 8, and generated the mixtures using a 3×3 mixing matrix. The proposed algorithm, the EASI and SOBI algorithm were then applied respectively to separate the mixtures. Note that, the step size in both the proposed algorithm and EASI is set to 0.002 in both tests. Each test was conducted three times for the separation of different mixtures generated by three different mixing matrices. Table I shows the final separation performance measured by the average of the performance indexes of all independent tests. It is clearly seen that the separation performance of the proposed periodic BSS algorithm is comparable to that of the EASI algorithm, which would break down with multiple Gaussian distributed sources. The final separation performance of SOBI is, of course, more accurate due to its block-based nature. However, this is in contrast to its computational complexity and latency due to the need to acquire a block of data.

The proposed algorithm is next used to address the fetal electrocardiogram (FECG) extraction problem. Fig. 9 shows the eight-channel cutaneous potential recording of a pregnant woman, obtained when eight electrodes are placed at different positions on the body of an expectant mother. The electrocardiogram measurements were recorded over 5 s, and sampled at 500 Hz[3] and represent mixtures of FECG and maternal electrocardiogram (MECG) contributions, as well as noise. The fetal heartbeat component, noise and respiratory motion artefacts are noticeable in the abdominal recordings, denoted in Fig. 9 as $x_1(t)$–$x_5(t)$, whereas FECG contributions are not visible in the thoracic measurements, which are dominated by the maternal heartbeat because of the distance between the foetus and the chest leads.

In order to apply BSS methods to FECG extraction, it has been shown in [28], [29] that the separation of MECG and
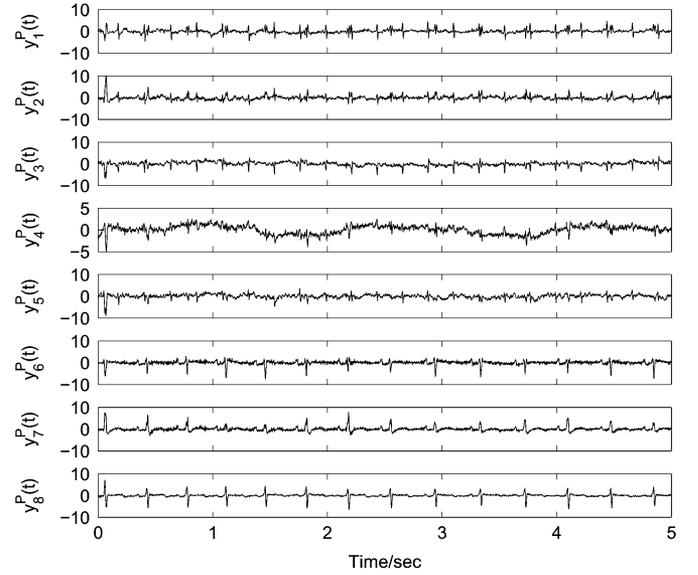
FECG components can be formulated as a BSS problem. The algorithm (38) was used to perform source separation, since the amplitude of the ECG measurements clearly varies rapidly, with periods of large and small magnitude alternating. Fig. 10 shows that the proposed algorithm extracts two fetal ($y_1^P(t)$ and $y_3^P(t)$) and three maternal ECG components ($y_6^P(t)$, $y_7^P(t)$, and $y_8^P(t)$), as well as extracting respiratory motion ($y_4^P(t)$) and other noise sources ($y_2^P(t)$ and $y_5^P(t)$).

D. Computational Load

One of the advantages of our proposed algorithm, as is convention for sequential algorithms, lies in its computational simplicity. To give a fair evaluation of the computational load of the proposed algorithm, we quantify its required operations (multiplications and additions), which are compared with those of the EASI and SOBI algorithms. It is worth noting that, due to the different strategies being taken for sequential algorithms (e.g., the proposed periodic BSS or EASI) and block algorithms (e.g., SOBI), it would be meaningful to compare the required (at least approximate) computations for approaching a similar separation quality (quantified by PI in this study), rather than to compare the required computations for each iteration in sequential algorithms with those in the whole block algorithms. Since the achievable final PI's are different for these approaches, as verified in Section V-C, it would be more appropriate to compare the required computations for achieving the stable separation by sequential algorithms, i.e., after their convergence, with those of the block algorithms. In Table II, both the general formulation of the required operations (denoted by the parameters given in previous sections) and also an example corresponding to Table I, are shown, where K is the number of time lags in the SOBI algorithm, and L is the length of the whole data sample, L^o is the number of the required data samples for the sequential algorithms to converge. Note that, in both test 1 and 2, L^o takes approximately 1300, $K = 7, L = 5000$. The ">" in Table II denotes the least operations required for SOBI. From Table II, it

TABLE II
COMPUTATIONAL COMPLEXITY OF THE PROPOSED ALGORITHM AS COMPARED WITH EASI AND SOBI ALGORITHMS

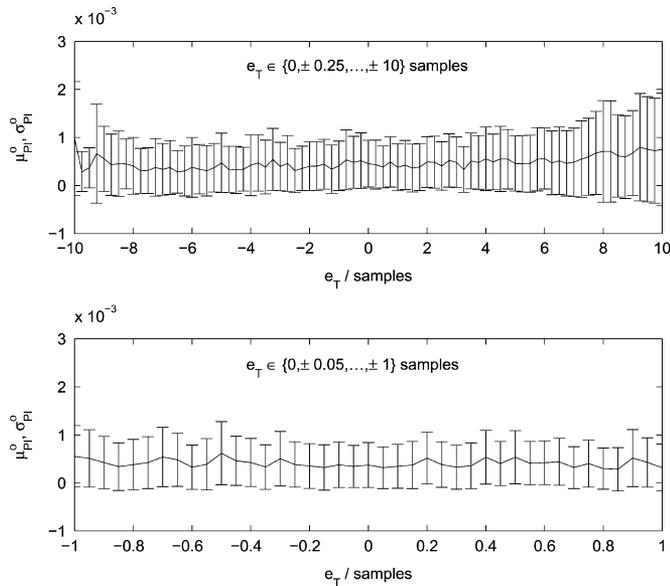| Algorithm	Periodic BSS using (29)	Periodic BSS using (35)	EASI	SOBI
Operations	$(2m^3 + 5m^2)L^o$	$(4m^3 + 4m^2 + 2m)L^o$	$(2m^3 + 7m^2)L^o$	$> (14+4K)m^3 + (2LK + 2L+1)m^2 + (1-2K)n - K$
Test 1	46800	67600	57200	> 320307
Test 2	128700	195000	152100	> 721097

Fig. 11. Behavior of the proposed algorithm as a function of the error in the selection of the lag τ. The plots show the mean and standard deviation of the PI value achieved at convergence, denoted as PI^o, over 100 independent trials. The upper plot shows the results obtained when an error varying between 0% to 10%, in steps of 0.25%, is present, while the lower plot illustrates performance when the error varies between 0% to 1%, in steps of 0.05%.

is observed that even when we count all the required iterations, the periodic BSS is still less complex than the SOBI algorithm, and computationally comparable to the EASI algorithm.

E. Robustness to Lag Selection

In this section we investigate the behavior of the proposed algorithm when the selected lag does not correspond to the exact fundamental period of one of the source signals. For the results presented here, two sinusoidal sources $s_1(k)$ and $s_2(k)$ were mixed by a real-time invariant mixing matrix, and zero mean white Gaussian noise was added such that the signal to noise ratio was 20 dB. The period of source $s_2(k)$ was fixed at $T_2 = 150$ samples, while the period of the first source was varied such that $(T_1 = 100 \pm e_T)$ samples, where e_T represents the error in the selection of the lag τ, which was fixed at $\tau = 100$ samples. Fig. 11 shows the average PI value, over 100 independent trials, and its standard deviation, achieved during separation, following convergence of the algorithm, denoted as PI^o. The upper plot illustrates the behavior of the proposed algorithm for $e_T \in \{0, \pm 0.25, \dots, \pm 10\}$ samples, corresponding to an error in the lag selection (error in the delay) varying between 0% and 10%, in steps of 0.25%. It shows that, in spite of

the increasing error, following convergence, the performance of the algorithm is fairly stable, leading to a PI value of approximately 5×10^{-4}, which increases to 7×10^{-4} when the error is roughly above 8%, and a standard deviation of approximately 5×10^{-4}, increasing to 7×10^{-4} for $e_T > 8$ samples. The lower plot shows the results obtained for an error varying between 0% and 1%, in steps of 0.05%. The low PI values and small fluctuations obtained following convergence in the simulations described, indicate that the algorithm is robust to errors in the selection of the lag. Since only a 10% error in either direction has been considered, an error tolerance of 10% has been demonstrated.

VI. CONCLUSION

In this paper a novel sequential algorithm for the blind separation of a class of periodic source signals is introduced, which exploits the temporal periodic nature of the source signals, and thus performs separation based only on second-order statistics. This algorithm is based on the use of the minimum period of the observed sources as a systematic method for lag selection when sequential blind source separation is performed based only on second-order statistics. Computer simulations have shown that periodic decorrelation allows the recovery of the original sources, and restoration of statistical independence. Moreover, Monte Carlo simulations have shown that the behavior of the proposed algorithm is comparable to that of EASI, during and following initial convergence. A normalized version of our proposed algorithm could easily be developed.

When several Gaussian source signals were present, it was found that the proposed algorithm extracted the periodic sources, while the EASI algorithm could not recover the original signals. The fetal ECG extraction problem was also addressed using the proposed method, which was found to separate the fetal and maternal ECG components, as well as extracting respiratory motion and other noise sources. Finally, it was shown that the algorithm is robust to errors in the selection of the lag. In particular, an error tolerance of approximately 10% was demonstrated. Furthermore, the proposed algorithm, in steady state, is shown to have similar separation performance to the EASI algorithm.

APPENDIX

A. Proof of Theorem 3

The proof follows from the fact that if two matrices M and N can be diagonalized, then they share the same eigenvector

matrix B if and only if $MN = NM$ [24]. Let $\mathbf{P}^o$ and $\mathbf{C}^o$ diagonalize $\mathbf{R}_y^o(T_i)$ and $\mathbf{R}_y^o(bT_i)$ respectively, for any integer b. Then

$$\mathbf{R}_y^o(T_i) = \mathbf{P}^o \mathbf{R}_s(T_i) \mathbf{P}^{oT} = \hat{\mathbf{D}}(T_i) \tag{42}$$

$$\mathbf{R}_y^o(bT_i) = \mathbf{C}^o \mathbf{R}_s(bT_i) \mathbf{C}^{oT} = \hat{\mathbf{D}}(bT_i). \tag{43}$$

Multiplying the two correlation matrices, we have

$$\mathbf{R}_y^o(T_i)\mathbf{R}_y^o(bT_i) = \hat{\mathbf{D}}(T_i)\hat{\mathbf{D}}(bT_i) \tag{44}$$

$$\mathbf{R}_y^o(bT_i)\mathbf{R}_y^o(T_i) = \hat{\mathbf{D}}(bT_i)\hat{\mathbf{D}}(T_i). \tag{45}$$

Both $\hat{\mathbf{D}}(bT_i)$ and $\hat{\mathbf{D}}(T_i)$ are diagonal; therefore we have $\hat{\mathbf{D}}(bT_i)\hat{\mathbf{D}}(T_i) = \hat{\mathbf{D}}(T_i)\hat{\mathbf{D}}(bT_i)$. Hence, $\mathbf{R}_y^o(T_i)$ and $\mathbf{R}_y^o(bT_i)$ share the same eigenvector matrix

$$\mathbf{R}_y^o(T_i)\mathbf{R}_y^o(bT_i) = \mathbf{P}^o \mathbf{R}_s(T_i)\mathbf{R}_s(bT_i)\mathbf{P}^{oT}. \tag{46}$$

Thus, consider the product of all output autocorrelation matrices $\prod_{b=-M}^{M} \mathbf{R}_y^o(bT_i)$, and the set of matrices $\mathbf{P}_b^o$ that diagonalize $\mathbf{R}_y^o(bT_i)$, such that in general

$$\prod_{b=-M}^{M} \mathbf{R}_y^o(bT_i) = \prod_{b=-M}^{M} \mathbf{P}_b^o \mathbf{R}_s(bT_i)\mathbf{P}_b^{oT}. \tag{47}$$

Since each autocorrelation matrix is diagonalized, the product of all matrices commutes. Expanding (47), gives

$$
\begin{aligned}
\prod_{b=-M}^{M} & \mathbf{R}_y^o(bT_i) \\
&= \prod_{b=-M}^{M} \mathbf{P}_b^o \mathbf{R}_s(bT_i)\mathbf{P}_b^{oT} \\
&= \mathbf{P}_{-M}^o \mathbf{R}_s(-MT_i)\mathbf{P}_M^{oT} \times \cdots \times \mathbf{P}_{-1}^o \mathbf{R}_s(-T_i)\mathbf{P}_{-1}^{oT} \\
&\quad \times \mathbf{P}_0^o \mathbf{R}_s(0)\mathbf{P}_0^{oT} \times \mathbf{P}_1^o \mathbf{R}_s(T_i)\mathbf{P}_1^{oT} \times \cdots \\
&\quad \times \mathbf{P}_M^o \mathbf{R}_s(MT_i)\mathbf{P}_M^{oT} \\
&= \hat{\mathbf{D}}(-MT_i) \times \cdots \times \hat{\mathbf{D}}(-T_i) \times \hat{\mathbf{D}}(0) \\
&\quad \times \hat{\mathbf{D}}(T_i) \times \cdots \times \hat{\mathbf{D}}(MT_i).
\end{aligned}
$$

From (42) and (43), we have

$$
\begin{aligned}
\prod_{b=-M}^{M} & \mathbf{R}_y^o(bT_i) \\
&= \mathbf{P}^o \mathbf{R}_s(-MT_i)\mathbf{P}^{oT} \times \cdots \times \mathbf{P}^o \mathbf{R}_s(-T_i)\mathbf{P}^{oT} \\
&\quad \times \mathbf{P}^o \mathbf{R}_s(0)\mathbf{P}^{oT} \times \mathbf{P}^o \mathbf{R}_s(T_i)\mathbf{P}^{oT} \times \cdots \\
&\quad \times \mathbf{P}^o \mathbf{R}_s(MT_i)\mathbf{P}^{oT} \\
&= \mathbf{P}^o[\mathbf{R}_s(-MT_i) \times \cdots \times \mathbf{R}_s(-T_i) \\
&\quad \times \mathbf{R}_s(0) \times \mathbf{R}_s(T_i) \times \cdots \times \mathbf{R}_s(MT_i)]\mathbf{P}^{oT}.
\end{aligned}
$$

Hence, diagonalizing $\mathbf{R}_y^o(T_i)$ simultaneously diagonalizes the output covariance matrix at every delay $\tau = nT_i, n \in \mathbb{Z}$.

B. Proof of Theorem 4

This can be shown by writing the output covariance matrix following convergence, in terms of $\mathbf{R}_s(T_i)$

$$
\begin{aligned}
\mathbf{R}_y^o(T_i)\mathbf{S}^o(\mathbf{y}(k,T_i)) &= \mathbf{P}^o \mathbf{D}(T_i)\mathbf{P}^{oH}\mathbf{S}^o(\mathbf{y}(k,T_i)) \\
&= \hat{\mathbf{D}}(T_i)\mathbf{S}^o(\mathbf{y}(k,T_i)) \\
&= \mathbf{\Lambda}_y
\end{aligned}
$$

where $\mathbf{S}^o(\mathbf{y}(k,T_i)) = \lim_{k \to \infty} \mathbf{S}(\mathbf{y}(k,T_i))$. Since $\hat{\mathbf{D}}(T_i)$ is the eigenvalue matrix of $\mathbf{R}_y^o(T_i)$, multiplication by $\mathbf{S}^o(\mathbf{y}(k,T_i))$ does not change its shape, but it simply changes the sign of the eigenvalues when they are negative. Hence, the eigenvalues of $\mathbf{R}_y^o(T_i)\mathbf{S}^o(\mathbf{y}(k,T_i))$ are the diagonal entries of $\mathbf{\Lambda}_y$, and are given by

$$\lambda_y^{(i)} = \hat{d}_{(ii)}(T_i)s_{ii} \tag{48}$$

where $\hat{d}_{(ii)}(T_i)$ and s_{ii} are, respectively, the diagonal elements of $\hat{\mathbf{D}}(T_i)$ and $\mathbf{S}^o(\mathbf{y}(k,T_i))$. Thus, from (32), (48) becomes

$$\lambda_y^{(i)} = \frac{\hat{d}_{(ii)}^2(T_i)}{|\hat{d}_{(ii)}(T_i)|} \tag{49}$$

which is always positive and real valued. It is straightforward that, during transient response, multiplication by the matrix $\mathbf{S}(\mathbf{y}(k,T_i))$ ensures that the diagonal elements of $\mathbf{R}_y(k,T_i)\mathbf{S}(\mathbf{y}(k,T_i))$ are always positive.

REFERENCES

[1] S. Amari and A. Cichocki, "Adaptive blind signal processing—Neural network approaches," *Proc. IEEE*, vol. 86, pp. 2026–2048, 1998.

[2] T. W. Lee, *Independent Component Analysis*. Norwell, MA: Kluwer, 1998.

[3] A. Mansour, A. K. Barros, and N. Ohnishi, "Blind separation of sources: Methods, assumptions and applications," *IEICE Trans. Fundam. Electron., Commun., Comput. Sci.*, vol. E83-A, pp. 1498–1512, 2000.

[4] S. Amari, A. Cichocki, and H. H. Yang, "A new learning algorithm for blind signal separation," in *Adv. Neural Inf. Process. Syst.*, vol. 8, 1996, pp. 752–763.

[5] C. Jutten and J. Herault, "Blind separation of sources, Part I: An adaptive algorithm based on a neuromimetic architecture," *Signal Process.*, vol. 24, pp. 1–10, 1991.

[6] P. Comon, "Independent component analysis, a new concept?," *Signal Process.*, vol. 36, pp. 287–314, 1994.

[7] A. J. Bell and T. J. Sejnowski, "An information-maximization approach to blind separation and blind deconvolution," *Neural Comput.*, vol. 7, pp. 1129–1159, 1995.

[8] J. F. Cardoso and B. Laheld, "Equivariant adaptive source separation," *IEEE Trans. Signal Process.*, vol. 44, no. 12, pp. 3017–3030, Dec. 1996.

[9] J. F. Cardoso and A. Souloumiac, "Blind beamforming for non-Gaussian signals," in *Proc. Inst. Elect. Eng.—F*, vol. 140, 1993, pp. 362–3370.

[10] K. Matsuoka, M. Ohya, and M. Kawamoto, "A neural net for blind separation of nonstationary signals," *Neural Netw.*, vol. 8, pp. 411–419, 1995.

[11] S. Choi and O. Y. Lee, "Nonstationary source separation," in *Proc. IEEE Region 10 Conf.*, vol. 1, 1999, pp. 670–673.

[12] D. L. Jones, "A new method for blind source separation of nonstationary signals," in *Proc. IEEE Int. Conf. Acoustic, Speech, Signal Processing (ICASSP)*, vol. 5, 1999, pp. 2893–2896.

[13] A. Ferreol and P. Chevalier, "On the behavior of current second and higher order blind source separation methods for cyclostationary sources," *IEEE Trans. Signal Process.*, vol. 48, no. 6, pp. 1712–1725, Jun. 2000.

[14] J. Antoni, J. Daniére, F. Guillet, and M. E. Badaoui, "Blind separation and identification of cyclostationary processes," in *Proc. IEEE Int. Conf. Acoustic, Speech, Signal Processing*, vol. 3, 2002, pp. 3077–3080.

[15] K. Abed-Meraim, Y. Xiang, J. H. Manton, and Y. Hua, "Blind source separation using second-order cyclostationary statistics," *IEEE Trans. Signal Process.*, vol. 49, no. 4, pp. 694–701, Apr. 2001.

[16] P. Chevalier, A. Ferreol, and L. Albera, "On the behavior of current second order blind source separation methods for first and second order cyclostationary sources—Application to CPFSK sources," in *Proc. IEEE Int. Conf. Acoustic, Speech, Signal Processing*, vol. 3, 2002, pp. 3081–3084.

[17] M. G. Jafari, S. R. Alty, and J. A. Chambers, "A new natural gradient algorithm for cyclostationary sources," in *Proc. Inst. Elect. Eng. Vision, Image, Signal Processing*, vol. 151, 2004, pp. 62–68.

[18] A. Belouchrani, K. Abed-Meraim, J. Cardoso, and E. Moulines, "A blind source separation technique using second-order statistics," *IEEE Trans. Signal Process.*, vol. 45, no. 2, pp. 434–444, Feb. 1997.

[19] S. Choi and A. Cichock, "Blind separation of nonstationary sources in noisy mixtures," *Electron. Lett.*, vol. 36, pp. 848–849, 2000.

[20] A. Belouchrani and A. Cichocki, "A robust procedure in blind source separation context," *Electron. Lett.*, vol. 36, pp. 2050–2051, 2000.

[21] A. K. Barros and N. Ohnishi, "Heart instantaneous frequency (HIF): An alternative approach to extract heart rate variability," *IEEE Trans. Biomed. Eng.*, vol. 48, no. 8, pp. 850–855, Aug. 2001.

[22] J. F. Cardoso, "Blind signal separation: statistical principles," *Proc. IEEE*, vol. 86, pp. 2009–2025, 1998.

[23] A. Belouchrani, K. Aded-Meraim, J. F. Cardoso, and E. Moulines, "Second-order blind separation of temporally correlated sources," in *Proc. Int. Conf. Digital Signal Processing*, 1993, pp. 346–351.

[24] G. Strang, *Linear Algebra and Its Applications*, 3rd ed. Orlando, FL: Harcourt Brace Jovanovich, 1988.

[25] A. Cichocki and S. Amari, *Adaptive Blind Signal and Image Processing: Learning Algorithms and Applications*. New York: Wiley, 2002.

[26] S. Amari, T. P. Chen, and A. Cichocki, "Nonholonomic orthogonal learning algorithms for blind sources separation," *Neural Comput.*, vol. 12, pp. 1463–1484, 2000.

[27] A. Cichocki, B. Orsier, A. Back, and S. Amari, "On-line adaptive algorithms in nonstationary environments using a modified conjugate gradient approach," in *Proc. IEEE Workshop Neural Networks for Signal Processing*, 1997, pp. 316–325.

[28] L. D. Lathauwer, B. D. Moor, and J. Vandewalle, "Fetal electrocardiogram extraction by blind source subspace separation," *IEEE Trans. Biomed. Eng.*, vol. 47, no. 5, pp. 567–572, May 2000.

[29] V. Zarzoso, A. K. Nandi, and E. Bacharakis, "Maternal and foetal ECG separation using blind source separation methods," *IMA J. Math. Appl. Med. Biol.*, vol. 14, pp. 207–225, 1997.

[30] A. Yeredor, "Non-orthogonal joint diagonalization in the least-squares sense with application in blind source separation," *IEEE Trans. Signal Process.*, vol. 50, no. 7, pp. 1545–1553, Jul. 2002.

[31] A. Cichocki and P. Georgiev, "Blind source separation algorithms with matrix constraints," *IEICE Trans. Fundam.*, vol. E86-A, no. 1, Jan. 2003.

[32] L. Parra and C. Spence, "Convolutive blind source separation of nonstationary sources," *IEEE Trans. Speech Audio Process.*, pp. 320–327, May 2000.

[33] K. Rahbar and J. Reilly, "Blind source separation algorithm for MIMO convolutive mixtures," in *Proc. Int. Workshop Independent Component Analysis Signal Separation*, Dec. 2001, pp. 242–247.

[34] E. Weinstein, M. Feder, and A. V. Oppenheim, "Multichannel signal separation by decorrelation," *IEEE Trans. Speech Audio Process.*, vol. 1, no. 4, pp. 405–413, Oct. 1993.

[35] S. Van-Gerven and D. Van-Compernolle, "Signal separation by symmetric adaptive decorrelation: Stability, convergence and uniqueness," *IEEE Trans. Signal Process.*, vol. 43, no. 7, pp. 1602–1612, Jul. 1995.

[36] Y. Xiang, K. Abed-Meraim, and Y. Hua, "Adaptive blind source separation by second order statistics and natural gradient," in *Proc. Int. Conf. Acoustic, Speech, Signal Processing (ICASSP)*, vol. 5, 1999, pp. 2917–2920.

[37] B. Xerri and B. Borloz, "An iterative method using conditional second-order statistics applied to the blind source separation problem," *IEEE Trans. Signal Process.*, vol. 52, no. 2, pp. 313–328, Feb. 2004.

[38] H. Buchner, R. Aichner, and W. Kellermann, "A generalization of blind source separation algorithms for convolutive mixtures based on second-order statistics," *IEEE Trans. Speech Audio Process.*, vol. 13, no. 1, pp. 120–134, Jan. 2005.

[39] A. Yeredor, "Blind separation of Gaussian sources via second-order statistics with asymptotically optimal weighting," *IEEE Signal Process. Lett.*, vol. 7, no. 7, pp. 197–200, Jul. 2000.

[40] B. Baykal, "Blind channel estimation via combining autocorrelation and blind phase estimation," *IEEE Trans. Circuits Syst. I*, vol. 51, no. 6, pp. 1125–1131, Jun. 2004.

[41] D. T. Pham, "Exploiting source non stationary and coloration in blind source separation," in *Proc. Int. Conf. Digital Signal Processing*, vol. 1, Jul. 2002, pp. 151–154.

Maria G. Jafari (S'01–M'02) was born in Rome, Italy. She received the M.Eng. degree in electrical and electronic engineering from Imperial College London, U.K., in 1999 and the Ph.D. degree in blind source separation from King's College London, U.K., in 2002.

From October 2002 to March 2004, she worked as a Research Associate at King's College London, where her research focused on the application of signal processing to biomedical problems. Currently, she is a Research Assistant with the Centre for Digital Music, Queen Mary University of London, London, U.K. Her research interests are in the areas of blind signal processing, speech and audio processing, and auditory perception.

Wenwu Wang (M'02) was born in Anhui, China, in 1974. He received the B.Sc. degree in automatic control, the M.E. degree in control science and control engineering, and the Ph.D. degree in navigation guidance and control, all from Harbin Engineering University, China, in 1997, 2000, and 2002, respectively.

In May 2002, he joined the Department of Electronic Engineering, King's College London, London, U.K., as a Postdoctoral Research Associate and transferred to the Cardiff School of Engineering, Cardiff University, Cardiff, Wales, U.K., in January 2004. In May 2005, he became a DSP Engineer with the Tao Group, Reading, Berkshire, U.K., where he worked on algorithm design and implementation for echo cancellation, dereverberation, and deconvolution and real-time and embedded system design for multimedia and graphic engines with application to mobile companies, computer companies, and consuming electronics companies. His current research interests are in the areas of blind signal processing, machine learning, and perception.

Dr. Wang is a member of the Institution for Electrical Engineers (IEE) and the IEEE Signal Processing and Circuits and Systems Societies.

Jonathon A. Chambers (S'85–M'85–SM'98) was born in Peterborough, U.K., in 1960. He received the B.Sc. (Hons.) degree from the Polytechnic of Central London, U.K., in 1985 and the Ph.D. degree from the University of London, U.K., in 1990, after study at Peterhouse, Cambridge University, Cambridge, U.K., and Imperial College London, U.K.

From 1979 to 1982, he served in the Royal Navy as an Artificer Apprentice in Action, Data, and Control. He has since held academic and industrial positions at Bath University, Imperial College London, King's College London, and Schlumberger Cambridge Research, Cambridge, U.K. In January 2004, he became a Cardiff Professorial Research Fellow within the Cardiff School of Engineering, Cardiff, Wales, U.K. He leads a team of researchers involved in the analysis, design, and evaluation of novel algorithms for digital signal processing with application in acoustics, biomedicine, and wireless communications. His research contributions have been in adaptive and blind signal processing. He has authored/coauthored close to 200 conference and journal publications and supervised 20 Ph.D. graduates.

Dr. Chambers is a past Chairman of the Institute for Electrical Engineers (IEE) Professional Group E5, Signal Processing. He has served as an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING and the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS, and he is currently serving as an Associate Editor for IEEE SIGNAL PROCESSING LETTERS and as a member of IEEE Technical Committee on Signal Processing and Methods. He received the Robert Mitchell Medal as the top graduate from Polytechnic of Central London.

Tetsuya Hoya (M'00) was born in Tokyo, Japan, on September 15, 1969. He received the B.Sc. and M.Sc. degrees in electrical engineering from Meiji University, Meiji, Japan, in 1992 and 1994, respectively and the Ph.D. degree from Imperial College of Science, Technology and Medicine, University of London, U.K., in 1998.

From April 1994 to September 1994, he was a Research Assistant with the Department of Electronics and Communication, Graduate School of Meiji University. From September 1997 to August 2000, he was a Postdoctoral Research Associate with the Department of Electrical and Electronics Engineering, Imperial College, London, U.K. Since October 2000, he has been a Research Scientist within the Brain Science Institute, RIKEN (The Institute of Physical and Chemical Research), Saitama, Japan, and, since April 2003, a visiting Lecturer at Saitama Institute Technology, Japan. His research interest focuses on a wide spectrum of brain science: artificial intelligence, cognitive neuroscience, combinatoric optimization, computational linguistics, consciousness studies, electroencephalography, neural networks (connectionism), philosophy, psychology, robotics, and signal processing. He has published more than 30 technical papers and is the author of the book *Artificial Mind System—Kernel Memory Approach* (New York: Springer-Verlag).

Dr. Hoya was a committee member of the 4th International Symposium on Independent Component Analysis and Blind Signal Separation (ICA 2003).

Andrzej Cichocki (M'96–SM'05) was born in Poland. He received the M.Sc. (Hons.), the Ph.D., and Habilitate Doctorate (Dr.Sc.) degrees, all in electrical engineering, from Warsaw University of Technology, Warsaw, Poland, in 1972, 1975, and 1982, respectively.

Since 1972, he has been with the Institute of Theory of Electrical Engineering and Electrical Measurements at the Warsaw University of Technology, where he became a full Professor in 1991. He is the coauthor of three books: *Adaptive Blind Signal and Image Processing* (New York: Wiley, 2003, with Prof. S. Amari); *MOS Switched-Capacitor and Continuous-Time Integrated Circuits and Systems* (New York: Springer, 1989); and *Neural Networks for Optimization and Signal Processing* (New York: Wiley and Teubner Verlag, 1993–1994, both with Prof. R. Unbehauen). He has also coauthored more than one 150 papers. Two of his books have been translated to Chinese and other languages. He spent a few years at the University Erlangen-Nuernberg (Germany) as Alexander Humboldt Research Fellow and Guest Professor working in the area of VLSI of electronic circuits, artificial neural networks, and optimization. He conducted and realized several successful research projects. From 1996 to 1999, he worked as a Team Leader of the Laboratory for Artificial Brain Systems, at the Frontier Research Program RIKEN, Japan, in the Brain Information Processing Group, and since 1999 he has been head of the laboratory for Advanced Brain Signal Processing in the Brain Science Institute, RIKEN, Saitama, Japan. His current research interests include biomedical signal and image processing (especially blind signal/image processing), neural networks and their applications, learning theory and robust algorithms, generalization and extensions of independent and principal component analysis, optimization problems and nonlinear circuits and systems theory and their applications.

Dr. Cichocki is a member of several international scientific committees and the Associate Editor of IEEE TRANSACTIONS ON NEURAL NETWORKS (since January 1998).