

BLIND IDENTIFICATION USING SECOND-ORDER STATISTICS: A NONSTATIONARITY AND NONWHITENESS APPROACH

Uttachai Manmontri and Patrick A. Naylor

Department of Electrical and Electronic Engineering, Imperial College London, SW7 2AZ, UK

Email: {uttachai.manmontri,p.naylor}@imperial.ac.uk

ABSTRACT

We consider an approach to the blind identification problem of instantaneous mixtures using second-order statistics through the nonstationarity and nonwhiteness properties of signals. We propose the use of natural gradient learning to form off-line/block processing (BP) and on-line processing (OP) algorithms suitable respectively for blind identification with batch data and on-line data and show that the proposed algorithms can be considered as a class of algorithms offering quasi-uniform performance. The identifiability conditions are presented which provide a key insight into these algorithms. The paper shows simulation results and concludes with some connections of the proposed algorithms to other existing algorithms.

1. INTRODUCTION

Blind identification of signals in an unknown environment using second-order statistics is being contemplated as an alternative to techniques employing higher-order statistics particularly when the signals are nonstationary and nonwhite. Several approaches based on the nonstationarity and nonwhiteness properties of signals have been proposed to solve the problem and most of them rely either explicitly or implicitly on the joint diagonalization criterion of correlation matrices. In [1], an on-line algorithm for blind identification of nonstationary signals using a logarithmic cost function is introduced. Such a cost function is considered to be related to the mutual information theory used in derivation of algorithms later proposed in [2]. A class of off-line algorithms which utilizes correlation matrices of nonstationary signals at different time intervals can be found in [2],[3] and the identifiability condition for a case of two correlation matrices of nonstationary signals is also given in [3]. An approach employing the nonwhiteness property of signals is proposed in [4] in which two correlation matrices at different time lags are used. A more general case using the nonwhiteness property at several time lags with an added whitening process is proposed in [5].

The combination of both properties are proposed for off-line processing of instantaneous mixtures in [6] and of convolutive mixtures in [7]. The technique of [6] exploits a whitening process followed by the elegant Jacobi-like technique to seek an orthogonal matrix that jointly diagonalizes several correlation matrices of whitened signals at different time intervals and different time lags. This method, although efficient, has been shown to introduce an imbalance of the weighting of the joint diagonalization by its whitening process [8] and additionally can be degraded by any residual additive noise that cannot be entirely removed by the whitening process [9]. The technique of [7] introduces a generalization of the logarithmic cost function proposed in [1] based

on a novel matrix formulation, also applicable for instantaneous mixtures, and uses a natural gradient adaptation to directly seek a nonorthogonal demixing matrix. Although it solves the unbalanced weighting of the joint diagonalization, the logarithmic cost function implies a need for source signals to be active throughout the process.

This paper uses second-order statistics through the nonstationarity and nonwhiteness properties of signals to solve the blind identification problem of instantaneous mixtures. The off-line/block processing (BP) and on-line processing (OP) algorithms are derived using natural gradient learning. We show that the proposed algorithms can be considered as a class of algorithms offering quasi-uniform performance. The BP algorithm does not require the signals to be active throughout the process. The identifiability conditions that provide a key insight into the algorithms are also given.

2. PROBLEM FORMULATION AND ASSUMPTIONS

We consider N source signals $\mathbf{s}(n) = [s_1(n), s_2(n), \dots, s_N(n)]^T$ and N observations of their instantaneous mixtures

$$\mathbf{x}(n) = \mathbf{A}\mathbf{s}(n) + \mathbf{v}(n) \quad (1)$$

where $\mathbf{v}(n)$ represents additive noise, \mathbf{A} is an $N \times N$ unknown mixing matrix and $n = 1, 2, \dots$ is the sample index. Given a set of the observed signals, our aim is to estimate the corresponding source signals within a scale factor and permutation ambiguity. We employ assumptions:

A1: \mathbf{A} is a square matrix with rank N .

A2: Each $s_i(n)$ is a zero mean, nonstationary and nonwhite process with $E[s_i(n)s_i(n-\tau)] \neq 0, \exists \tau \neq 0, \forall i = 1, 2, \dots, N$, where τ is the time lag, $E[\cdot]$ denotes the expectation operator and \exists denotes *for some*. Also, each $s_i(n)$ is uncorrelated with $E[s_i(n)s_j(n-\tau)] = E[s_i(n)]E[s_j(n-\tau)], \forall \tau, \forall 1 \leq i \neq j \leq N$.

A1 ensures the existence of all source signals to be observed in the form of $\mathbf{x}(n)$ by the rank of \mathbf{A} . A2 is a key assumption for our proposed algorithms.

3. OFF-LINE/BLOCK PROCESSING (BP) ALGORITHM

The off-line/block processing (BP) algorithm uses several blocks of observed signals indexed $k = 1, 2, \dots, K$. We define a block which is sufficiently small to capture a set of observed signals that exhibits wide-sense stationarity giving

$$\mathbf{X}_k(\bar{n}) = [\mathbf{x}(L_{k-1} - D_k + 1), \dots, \mathbf{x}(L_{k-1} - D_k + L_k)] \quad (2)$$

where $\mathbf{X}_k(\bar{n})$ is an $N \times L_k$ observed signal matrix, L_k is the block length with $L_0 = 0$, D_k is the overlap factor with $0 \leq D_k < L_k$ and $D_1 = 0$, and \bar{n} represents the sample index corresponding to the start of the block. It is important that (2) has zero mean. This can be achieved by subtracting the mean of the observed signals in each block.

In a *noiseless* case, we can formulate the relation between the observed signals and source signals in a form of *local* time-average correlation matrices using (1), (2) and A2 as

$$\mathbf{R}_x^{(k,\tau)} = \mathbf{A} \Lambda_s^{(k,\tau)} \mathbf{A}^T, \forall k = 1, 2, \dots, K, \forall \tau = 0, 1, \dots, \Gamma \quad (3)$$

where Γ represents the maximum nonzero time lag of interest, $\mathbf{R}_x^{(k,\tau)} = \frac{1}{L_k} \mathbf{X}_k(\bar{n}) \mathbf{X}_k^T(\bar{n} - \tau)$ denotes the local time-average correlation matrix which is averaged across the observed signals at block and time lag (k, τ) and, similarly, $\Lambda_s^{(k,\tau)}$ denotes the local time-average diagonal correlation matrix of source signals with its corresponding diagonal elements $r_{s_i}^{(k,\tau)}$, for $i = 1, 2, \dots, N$.

Before proceeding further, since we assume the nonstationarity and nonwhiteness of source signals (A2), it is useful to say how nonstationary and nonwhite the source signals should be for us to blindly identify them. The following identifiability conditions give necessary and sufficient requirements for our algorithm.

Identifiability conditions. (i) Blind identification of nonstationary and nonwhite sources is possible using the *nonstationarity* as a primary property together with the local time-average correlation matrices of the observed signals iff $\mathbf{r}_{s_i}^{(\tau)} = [r_{s_i}^{(1,\tau)}, r_{s_i}^{(2,\tau)}, \dots, r_{s_i}^{(K,\tau)}]$ and $\mathbf{r}_{s_j}^{(\tau)} = [r_{s_j}^{(1,\tau)}, r_{s_j}^{(2,\tau)}, \dots, r_{s_j}^{(K,\tau)}]$ are linearly independent, $\exists \tau, \forall 1 \leq i \neq j \leq N$; (ii) blind identification of nonstationary and nonwhite sources is possible using the *nonwhiteness* as a primary property together with the local time-average correlation matrices of the observed signals iff $\tilde{\mathbf{r}}_{s_i}^{(k)} = [r_{s_i}^{(k,0)}, r_{s_i}^{(k,1)}, \dots, r_{s_i}^{(k,\Gamma)}]$ and $\tilde{\mathbf{r}}_{s_j}^{(k)} = [r_{s_j}^{(k,0)}, r_{s_j}^{(k,1)}, \dots, r_{s_j}^{(k,\Gamma)}]$ are linearly independent, $\exists k, \forall 1 \leq i \neq j \leq N$.

Proof: We prove here that the conditions (i) and (ii) are necessary. (i) Consider two unknown mixing matrices $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_N]$ and $\tilde{\mathbf{A}} = [\tilde{\mathbf{a}}_1, \tilde{\mathbf{a}}_2, \dots, \tilde{\mathbf{a}}_N]$ with $[\mathbf{a}_1, \mathbf{a}_2]$ and $[\tilde{\mathbf{a}}_1, \tilde{\mathbf{a}}_2]$ defined by

$$[\tilde{\mathbf{a}}_1, \tilde{\mathbf{a}}_2] = [\mathbf{a}_1, \mathbf{a}_2] \begin{bmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{bmatrix}$$

where \mathbf{a}_i and $\tilde{\mathbf{a}}_i$, $i = 1, 2, \dots, N$, are $N \times 1$ vectors. Also, consider two $N \times 1$ source signal vectors $\mathbf{s}(n)$ and $\tilde{\mathbf{s}}(n)$ defined in the same fashion with $[s_1(n), s_2(n)]^T$ and $[\tilde{s}_1(n), \tilde{s}_2(n)]^T$ given by

$$\begin{bmatrix} \tilde{s}_1(n) \\ \tilde{s}_2(n) \end{bmatrix} = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix} \begin{bmatrix} s_1(n) \\ s_2(n) \end{bmatrix}.$$

Let us assume $\mathbf{r}_{s_1}^{(\tau)} = \mathbf{r}_{s_2}^{(\tau)}$, then we readily see that $\mathbf{x}(n) = \mathbf{A}\mathbf{s}(n)$ and $\tilde{\mathbf{x}}(n) = \tilde{\mathbf{A}}\tilde{\mathbf{s}}(n)$ as well as $\mathbf{s}(n)$ and $\tilde{\mathbf{s}}(n)$ share the same local time-average correlation matrices. Hence, we cannot identify these signals by using the local time-average correlation matrices. In addition, this condition generalizes the condition found in proposition 1 of [3], where two correlation matrices of nonstationary signals at two different time intervals are considered.

(ii) This condition can be viewed as a generalization of the condition on the normalized spectra of source signals in [5], which utilizes power normalization by considering $r_{s_i}^{(\kappa,0)} = r_{s_j}^{(\kappa,0)} = 1$, where κ denotes the whole set of samples, and as a special case of the condition found in [10], when a cycle frequency of cyclostationary source signals i and j is zero. In addition, a slightly different condition with proof for complex signals is given in [11].

Even though the results in [5],[11] and [10] are for complete signals whereas we are concerned with block-based processing, we nevertheless propose that the proof can similarly be applied for appropriate choice of L_k .

In reality, the source signals are not accessible nor are their nonstationarity and nonwhiteness known a priori. The BP algorithm requires only that $\mathbf{r}_{s_i}^{(\tau)}$ is linearly independent of $\mathbf{r}_{s_j}^{(\tau)}$, $\exists \tau$ or $\tilde{\mathbf{r}}_{s_i}^{(k)}$ is linearly independent of $\tilde{\mathbf{r}}_{s_j}^{(k)}$, $\exists k$, for any set of local time-averaged observed signal correlation matrices whether or not the matrices are zero or have similar statistics.

Let $\mathbf{y}(n) = \mathbf{W}\mathbf{x}(n)$ be an $N \times 1$ vector of output signals, where \mathbf{W} is an $N \times N$ demixing matrix giving $\mathbf{W}\mathbf{A} = \mathbf{C}$, where \mathbf{C} is a global system matrix. We premultiply and postmultiply (3) with \mathbf{W} and its transpose to get

$$\Lambda_s^{(k,\tau)} \doteq \mathbf{W}\mathbf{R}_x^{(k,\tau)}\mathbf{W}^T = \mathbf{R}_y^{(k,\tau)} \quad (4)$$

where $\mathbf{R}_y^{(k,\tau)}$ is the local time-average correlation matrix of the output signals at (k, τ) and \doteq denotes an equality up to an arbitrary scale and permutation which are represented within \mathbf{C} . It is clear from (4) that, to recover the source signals, the local time-average output correlation matrix $\mathbf{R}_y^{(k,\tau)}$ must be diagonal, representing the statistics of source signals up to their arbitrary scaling and permutation.

In view of the identifiability conditions and (4), two local time-average correlation matrices of source signals satisfying either (i) or (ii) provide adequate conditions to solve the problem through their corresponding observed signal correlation matrices. However, a local time-average correlation matrix is, in fact, an *estimate* of the ensemble average correlation matrix from available samples and thus does not provide the exact measurement. Accordingly, the more conditions of (4) (via different (k, τ)) that the demixing matrix satisfies, the more accurate the solution of the problem will become (see e.g. [7] and references therein).

To obtain a nonorthogonal demixing matrix \mathbf{W} as described in [8] and to obtain a more compact form of gradient adaptation presented thereafter, we use the symmetric part of local time-average correlation matrices and rewrite (4) as

$$\bar{\Lambda}_s^{(k,\tau)} \doteq \mathbf{W}\bar{\mathbf{R}}_x^{(k,\tau)}\mathbf{W}^T = \bar{\mathbf{R}}_y^{(k,\tau)} \quad (5)$$

where $\bar{\mathbf{R}}_x^{(k,\tau)}$, $\bar{\Lambda}_s^{(k,\tau)}$ and $\bar{\mathbf{R}}_y^{(k,\tau)}$ are respectively the symmetric part of $\mathbf{R}_x^{(k,\tau)}$, $\Lambda_s^{(k,\tau)}$ and $\mathbf{R}_y^{(k,\tau)}$ such that e.g. $\bar{\mathbf{R}}_x^{(k,\tau)} = \frac{1}{2}(\mathbf{R}_x^{(k,\tau)} + \mathbf{R}_x^{T(k,\tau)})$.

We now introduce the following joint diagonalization cost function to be minimized as

$$J_D(\mathbf{W}) = \sum_{k=1}^K \sum_{\tau=0}^{\Gamma} \beta^{(k,\tau)} \left\| \bar{\mathbf{R}}_y^{(k,\tau)} - \text{diag}(\bar{\mathbf{R}}_y^{(k,\tau)}) \right\|_F^2 \quad (6)$$

where $\beta^{(k,\tau)}$ is a positive weight satisfying $\sum_{k=1}^K \sum_{\tau=0}^{\Gamma} \beta^{(k,\tau)} = 1$ and is generally set to $\frac{1}{K(\Gamma+1)}$ giving balanced weighting for joint diagonalization, $\text{diag}(\cdot)$ is the operator that zeros off-diagonal elements and $\|\cdot\|_F$ denotes the Frobenius norm. The following theorem provides the existence of a stationary point of J_D .

Theorem. Assume, without loss of generality, that the identifiability conditions holds and \mathbf{W} is rank N , then J_D has a stationary point iff \mathbf{W} is a demixing matrix.

Proof: Let $\nabla_{\mathbf{W}}$ be a gradient operator with respect to \mathbf{W} , we have

$$\nabla_{\mathbf{W}} J_D = 4 \sum_{k=1}^K \sum_{\tau=0}^{\Gamma} \beta^{(k,\tau)} \left(\bar{\mathbf{R}}_y^{(k,\tau)} - \text{diag}(\bar{\mathbf{R}}_y^{(k,\tau)}) \right) \mathbf{W} \bar{\mathbf{R}}_x^{(k,\tau)}.$$

If \mathbf{W} is a demixing matrix, it can be shown that $\nabla_{\mathbf{W}} J_D = 0$.

Conversely, since $\mathbf{W} \neq \mathbf{0}$ because it is rank N , $\hat{\mathbf{R}}_{\mathbf{x}}^{(k,\tau)} \neq \mathbf{0}$, $\exists k$ and $\exists \tau$ using the identifiability conditions and A1, it follows that $\nabla_{\mathbf{W}} J_D = 0$, if \mathbf{W} is a demixing matrix.

Since no closed form solution exists for this problem, we seek a demixing matrix by employing natural gradient learning (Theorem 6 of [12]). In addition, by applying natural gradient learning to J_D , the resulting algorithm is equivariant, meaning that it exhibits uniform performance in such a way that its behavior is independent of the mixing matrix [13]. The natural gradient of J_D is given by

$$\tilde{\nabla}_{\mathbf{W}} J_D = 4 \sum_{k=1}^K \sum_{\tau=0}^{\Gamma} \beta^{(k,\tau)} \left(\hat{\mathbf{R}}_{\mathbf{y}}^{(k,\tau)} - \text{diag} \left(\hat{\mathbf{R}}_{\mathbf{y}}^{(k,\tau)} \right) \right) \hat{\mathbf{R}}_{\mathbf{y}}^{(k,\tau)} \mathbf{W} \quad (7)$$

where $\tilde{\nabla}_{\mathbf{W}}$ is the natural gradient operator with respect to \mathbf{W} .

Notice that we assume \mathbf{W} to have rank N , which is the direct consequence of A1. As mentioned in the previous Section, the rank N of \mathbf{A} results in the existence of every source signal in a form of observed signals when all source signals are sufficiently rich to excite (1). Based on this point, some algorithms, without resorting to constraining \mathbf{W} , exploit a logarithmic function of an output correlation matrix [1] or of local time-average output correlation matrices [7] as their cost function to identify \mathbf{W} . This method requires that such a matrix or matrices are positive definite so that they are invertible. Moreover, if the source signals are very weak, gradient noise amplification can be a problem. In other words, this implies the need for permanent excitation of sources in every signal block. As an alternative to the requirement for permanent excitation, \mathbf{W} can be constrained to have rank N . However, this would normally degrade algorithm performance in a noisy environment [9]. In our work, we solve this problem by proposing to minimize an unconstrained cost function but with the addition of a quadratic penalty function, J_P , that constrains all diagonal elements of \mathbf{W} to be unity

$$J_{BP}(\mathbf{W}) = J_D + \lambda_P J_P \quad (8)$$

where $J_P = \|\text{diag}(\mathbf{W} - \mathbf{I})\|_F^2$, \mathbf{I} is the $N \times N$ identity matrix and λ_P is a small positive constant. From optimization theory, as $\lambda_P \rightarrow 0$, ($\lambda_P \neq 0$), \mathbf{W} will adapt in a more flexible way rather than strictly constraining its diagonal elements, nonetheless, still keeping \mathbf{W} to have rank N . The natural gradient of J_P is given by

$$\tilde{\nabla}_{\mathbf{W}} J_P = 2 \text{diag}(\mathbf{W} - \mathbf{I}) \mathbf{W}^T \mathbf{W}. \quad (9)$$

Accordingly, the BP algorithm updates the demixing matrix in an iterative fashion using

$$\mathbf{W}(m+1) = \mathbf{W}(m) - \mu \left(\tilde{\nabla}_{\mathbf{W}} J_D(m) + \lambda_P \tilde{\nabla}_{\mathbf{W}} J_P(m) \right) \quad (10)$$

where μ is a positive step size and m is the iterative index. It can be seen that (10) violates the uniform performance property due to $\tilde{\nabla}_{\mathbf{W}} J_P$. However, we regard the BP algorithm as an algorithm that exhibits quasi-uniform performance since λ_P is always much less than unity and therefore considerably reduces the effect of $\tilde{\nabla}_{\mathbf{W}} J_P$.

4. ON-LINE PROCESSING (OP) ALGORITHM

The on-line processing (OP) algorithm can be viewed as a modified BP algorithm. More specifically, each block of signals consists of data from the beginning of the process to the current observed sample and therefore these blocks grow in size from block

to block. Using this methodology, the number of blocks is equal to the number of samples. Replacing the block index k with the sample index n , we calculate a *current* time-average correlation matrix $\hat{\mathbf{R}}_{\mathbf{x}}^{(n,\tau)}$ by deemphasizing the past observed signals using the following nonparametric recursive relation

$$\hat{\mathbf{R}}_{\mathbf{x}}^{(n,\tau)} = \alpha \hat{\mathbf{R}}_{\mathbf{x}}^{(n-1,\tau)} + (1 - \alpha) \mathbf{x}(n) \mathbf{x}^T(n - \tau) \quad (11)$$

where α is the fading factor with $0 \leq \alpha \leq 1$. If the source signals are stationary, (11) consequently reduces to the sample average simply calculated by

$$\hat{\mathbf{R}}_{\mathbf{x}}^{(n,\tau)} = \frac{n-1}{n} \hat{\mathbf{R}}_{\mathbf{x}}^{(n-1,\tau)} + \frac{1}{n} \mathbf{x}(n) \mathbf{x}^T(n - \tau). \quad (12)$$

Similarly, let $\hat{\Lambda}_{\mathbf{s}}^{(n,\tau)}$ be a current time-average diagonal correlation matrix of source signals with its corresponding diagonal elements $r_{s_i}^{(n,\tau)}$, for $i = 1, 2, \dots, N$. We have the identifiability conditions for the OP algorithm as (i) $\mathbf{r}_{s_i}^{(\tau)} = [r_{s_i}^{(1,\tau)}, r_{s_i}^{(2,\tau)}, \dots]$ is linearly independent of $\mathbf{r}_{s_j}^{(\tau)} = [r_{s_j}^{(1,\tau)}, r_{s_j}^{(2,\tau)}, \dots]$, $\exists \tau$; (ii) $\hat{\mathbf{r}}_{s_i}^{(n)} = [r_{s_i}^{(n,0)}, r_{s_i}^{(n,1)}, \dots, r_{s_i}^{(n,\Gamma)}]$ is linearly independent of $\hat{\mathbf{r}}_{s_j}^{(n)} = [r_{s_j}^{(n,0)}, r_{s_j}^{(n,1)}, \dots, r_{s_j}^{(n,\Gamma)}]$. It is worth noting that if only (i) holds at $\tau = 0$, this leads to the concept underlying the on-line algorithms in [1],[2], which utilize only the nonstationarity of signals, and also reveals that it is not the nonstationarity of signals that enables signal identification but the difference of nonstationarity between signals.

The unconstrained cost function for OP is given by

$$J_{OP}(n) = J_D(n) + \lambda_P J_P(n) \quad (13)$$

where $J_D(n) = \sum_{\tau=0}^{\Gamma} \beta^{(n,\tau)} \left\| \hat{\mathbf{R}}_{\mathbf{y}}^{(n,\tau)} - \text{diag} \left(\hat{\mathbf{R}}_{\mathbf{y}}^{(n,\tau)} \right) \right\|_F^2$,

$J_P(n) = \|\text{diag}(\mathbf{W}(n) - \mathbf{I})\|_F^2$, $\hat{\mathbf{R}}_{\mathbf{y}}^{(n,\tau)}$ is the symmetric part of a current time-average output correlation matrix and $\sum_{\tau=0}^{\Gamma} \beta^{(n,\tau)} = 1$. In the general case, $\beta^{(n,\tau)}$ is set to $\frac{1}{\Gamma+1}$.

The natural gradient is given by

$$\tilde{\nabla}_{\mathbf{W}} J_D(n) = 4 \sum_{\tau=0}^{\Gamma} \left(\hat{\mathbf{R}}_{\mathbf{y}}^{(n,\tau)} - \text{diag} \left(\hat{\mathbf{R}}_{\mathbf{y}}^{(n,\tau)} \right) \right) \hat{\mathbf{R}}_{\mathbf{y}}^{(n,\tau)} \mathbf{W}(n) \quad (14)$$

and the update equation of the OP algorithm takes the form

$$\mathbf{W}(n+1) = \mathbf{W}(n) - \mu \left(\tilde{\nabla}_{\mathbf{W}} J_D(n) + \lambda_P \tilde{\nabla}_{\mathbf{W}} J_P(n) \right) \quad (15)$$

where $\tilde{\nabla}_{\mathbf{W}} J_P(n)$ can be obtained from (9). For the OP algorithm to attain its minimum, the source signals must persistently excite (1).

5. SIMULATIONS

We have evaluated the performance of the proposed algorithms by means of the performance index (PI) = $\frac{1}{2N(N-1)} \left(\sum_{i=1}^N \left(\sum_{j=1}^N \frac{c_{ij}^2}{\max_j(c_{ij}^2)} - 1 \right) + \sum_{j=1}^N \left(\sum_{i=1}^N \frac{c_{ij}^2}{\max_i(c_{ij}^2)} - 1 \right) \right)$, where c_{ij} is an element in the i^{th} row and j^{th} column of the global system matrix \mathbf{C} and $\max_{i,j}(\cdot)$ is the maximum value for $1 \leq i, j \leq N$. Smaller values of PI indicate better performance.

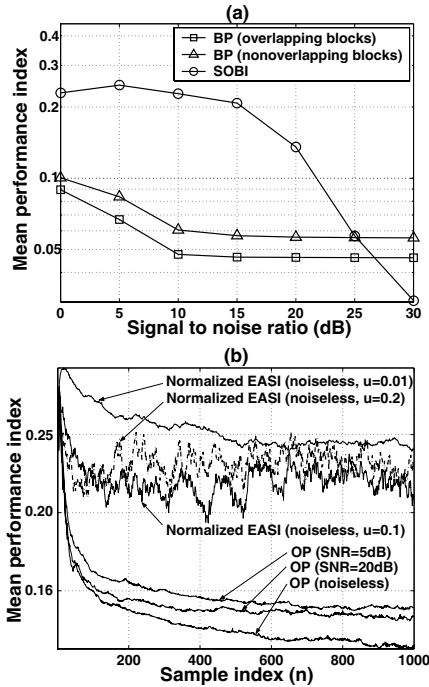


Fig. 1. (a) Mean performance indices of the BP and SOBI algorithms obtained from the mixtures of two speech signals. (b) Mean performance indices of the OP and normalized EASI algorithms obtained from the mixtures of two stationary but nonwhite signals generated by AR(2) and AR(4) models.

In the first simulation, two speech signals¹ are used as sources associated with noise from two autoregressive models of order four (AR(4)). The blind identification is performed at various levels of signal to noise ratio (SNR) by the BP algorithm with overlapping and nonoverlapping block methods and, for comparison, by the second-order blind identification (SOBI) algorithm² [5], whose performance in a noisy environment depends on how well a whitening process deals with noise. The PI is averaged over 10 independent trials of the noise generating processes and randomly chosen mixing matrices. In Fig.1(a), we can see that both BP algorithms outperform SOBI except at high SNR and superior performance is obtained using overlapping blocks.

In the second simulation, we employ the OP algorithm to identify two stationary but nonwhite sources from AR(2) and AR(4) models associated with noise from two AR(1) models and average it over 500 independent trials of all signals generating processes and randomly chosen mixing matrices. The results are compared with the normalized equivariant adaptive source separation via independence (EASI) algorithm [13], which is based on higher-order statistics. As expected, Fig.1(b) shows improved performance of the OP algorithms over that of the normalized EASI.

6. CONCLUSIONS AND DISCUSSION

We have presented the BP and the OP blind signal identification algorithms which employ second-order signal statistics and non-

stationarity and nonwhiteness properties. By using natural gradient learning, the proposed algorithms exhibit quasi-uniform performance such that their behavior is almost independent of the mixing matrix. Identifiability conditions have been given.

Our BP algorithm differs from [14],[15] because of the use of nonstationarity and nonwhiteness properties of signals and the introduction of natural gradient learning, even though they are all based on joint diagonalization of correlation matrices. In contrast to [7], the overlapping block method is introduced and the implicit requirement for persistently active excitation of sources in every block is avoided in our technique. Except at high SNR, results of our tests show significant improvement in performance compared to SOBI, especially when overlapping blocks are employed. The OP algorithm includes the concept that is employed in [1],[2] but extends it to consider both nonstationarity and nonwhiteness properties of signals.

7. REFERENCES

- [1] K. Matsuo, M. Ohya, and M. Kawamoto, "A neural net for blind separation of nonstationary signals," *Neural Networks*, vol. 8, pp. 411–419, 1995.
- [2] D.-T. Pham and J.-F. Cardoso, "Blind separation of instantaneous mixtures of nonstationary sources," *IEEE Trans. Signal Processing*, vol. 49, pp. 1837–1848, Sept. 2001.
- [3] M. K. Tsatsanis and C. Kweon, "Blind source separation of nonstationary sources using second-order statistics," in *Proc. 32nd Asilomar Conference on Signals, Systems and Computers*, CA, USA, Nov. 1998, pp. 1574–1578.
- [4] L. Molgedey and H. G. Schuster, "Separation of a mixture of independent signals using time delayed correlations," *Physical Review Letters*, vol. 72, pp. 3634–3637, June 1994.
- [5] A. Belouchrani, K. Abed-Meraim, J.-F. Cardoso, and E. Moulines, "A blind source separation technique using second-order statistics," *IEEE Trans. Signal Processing*, vol. 45, pp. 434–444, Feb. 1997.
- [6] S. Choi and A. Cichocki, "Blind separation of nonstationary sources in noisy mixtures," *Electronics Letters*, vol. 36, no. 9, pp. 848–849, Apr. 2000.
- [7] H. Buchner, R. Aichner, and W. Kellermann, "A generalization of a class of blind source separation algorithms for convolutive mixtures," in *Proc. ICA'03*, Nara, Japan, Apr. 2003, pp. 945–950.
- [8] A. Yeredor, "Non-orthogonal joint diagonalization in the least-squares sense with application in blind source separation," *IEEE Trans. Signal Processing*, vol. 50, pp. 1545–1553, July 2002.
- [9] J.-F. Cardoso, "On the performance of orthogonal source separation algorithms," in *Proc. EUSIPCO'94*, Edinburgh, UK, Sept. 1994, pp. 776–779.
- [10] K. Abed-Meraim, Y. Xiang, J. H. Manton, and Y. Hua, "Blind source-separation using second-order cyclostationary statistics," *IEEE Trans. Signal Processing*, vol. 49, pp. 694–701, Apr. 2001.
- [11] K. Abed-Meraim, Y. Xiang, and Y. Hua, "Generalized second order identifiability condition and relevant testing technique," in *Proc. IEEE ICASSP'00*, Istanbul, Turkey, June 2000, pp. 2989–2992.
- [12] S. Amari, "Natural gradient works efficiently in learning," *Neural Computation*, vol. 10, pp. 251–276, Feb. 1998.
- [13] J.-F. Cardoso and B. H. Laheld, "Equivariant adaptive source separation," *IEEE Trans. Signal Processing*, vol. 44, pp. 3017–3030, Dec. 1996.
- [14] L. Parra and C. Spence, "Convolutional blind separation of nonstationary sources," *IEEE Trans. Speech Audio Processing*, vol. 8, pp. 320–327, May 2000.
- [15] M. Joho and H. Mathis, "Joint diagonalization of correlation matrices by using gradient methods with application to blind signal separation," in *Proc. IEEE SAM'02*, VA, USA, Aug. 2002, pp. 273–277.

¹From "http://www.bsp.brain.riken.jp/ICALAB/ICALABSignalProc/benchmarks."

²SOBI attains a superior performance when signals are stationary.