MIMO-AR SYSTEM IDENTIFICATION AND BLIND SOURCE SEPARATION USING GMM

Tirza Routtenberg and Joseph Tabrikian

Dept. of Electrical and Computer Engineering Ben-Gurion University of the Negev, Beer-Sheva 84105, Israel

ABSTRACT

The problem of blind source separation (BSS) for multiple-input multiple-output (MIMO) autoregressive (AR) mixtures is addressed in this paper. A new time-domain method for system identification and BSS is proposed based on the Gaussian mixture model (GMM) for sources distribution. The algorithm is based on the generalized expectation-maximization (GEM) method for joint estimation of the AR model parameters and the GMM parameters of the sources. The method is tested via simulations of synthetic and real audio signals. The results show that the proposed algorithm outperforms the well-known multidimensional linear predictive coding (LPC), and it achieves higher signal-to-interference ratio (SIR) in the BSS problem.

Index Terms— MIMO-AR, MIMO system identification, BSS, EM, convolutive mixtures, GMM

1. INTRODUCTION

Blind source separation (BSS) has been intensively investigated in the literature in the recent two decades. This problem is important in several applications in communications, biomedical engineering and blind audio sources separation. Classical research in this field has dealt with blind separation of instantaneous mixtures where the observed signal samples are obtained by a linear combination of the sources samples [1], [2]. A more challenging problem is multipleinput multiple-output (MIMO) system identification and blind separation of convolutive mixtures where the mixing matrix is frequencydependent. There are two common approaches to handle the convolutive BSS problem: frequency domain [3], [4], and time domain [5], [6]. In the frequency domain approach, the sensor signal model at each frequency bin is similar to the instantaneous mixture case. Therefore, instantaneous BSS techniques can be independently applied for each frequency bin. The main disadvantage of this approach is the frequency permutations problem [4].

In this paper, the time domain approach is adopted for the convolutive BSS problem. The convolutive mixture is modeled by a MIMO system with multi-dimensional autoregressive (AR) relations between the inputs and outputs. The sources distribution is modeled by the Gaussian mixture model (GMM). The motivation for modeling the non-Gaussian source signals by GMM is that many probability density functions (PDF) can be closely approximated by a mixture of finite number of Gaussians [7]. The GMM distribution combined with the AR model for the source signal has been used for BSS of instantaneous mixtures [8].

In this work, we consider a different problem with independent and identically distributed (i.i.d.) GMM sources and convolutive MIMO-AR mixtures. We consider two problems: system identification and source separation. The MIMO-AR parameters of the mixture and the GMM parameters are jointly estimated via a quasi maximum-likelihood (QML) estimator, implemented by using the generalized expectation-maximization (GEM) iterative method. The proposed algorithm is a generalization of the single-input singleoutput (SISO) system identification problem presented in [9], which is an extension of the Yule Walker equations to GMM.

2. PROBLEM FORMULATION

2.1. The MIMO-AR model

Consider the following MIMO-AR model:

$$\mathbf{x}_n = \mathbf{A}\mathbf{x}_n^P + \mathbf{H}\mathbf{s}_n \quad \forall n = 1, \dots, N \tag{1}$$

where *n* represents the time index, \mathbf{s}_n denotes the *K*-dimensional source vector, $\mathbf{x}_n = (x_{n,1}, \ldots, x_{n,L})^T$ is an *L*-dimensional observation vector, and $x_{n,i}$ represents the *n*th sample at the *i*th sensor. The past samples vector, \mathbf{x}_n^P , is defined as follows

$$\mathbf{x}_{n}^{P} = (x_{n-1,1}, \dots, x_{n-P,1}, \dots, x_{n-1,L}, \dots, x_{n-P,L})^{T}$$

where P is the AR order, assumed to be known. A is an unknown $L \times LP$ deterministic state transition matrix relating the state vector of the system between time instance n and time instances $n - 1, \ldots, n - P$. The matrix H is an unknown $L \times K$ deterministic input mixing matrix. The number of sensors, L, is assumed to be greater or equal to the number of sources, K. This assumption is not necessary for estimation of A. In addition, we assume that the initial conditions are known.

2.2. Source distribution model

The GMM is commonly used to model non-Gaussian PDF's, since it is capable of closely approximating many densities and has been considered by a number of researchers for this purpose [7]. In this work, the sources are assumed to be statistically independent. Each source signal is an i.i.d. GMM-distributed sequence. Under this model, the PDF of the *k*th source signal at each time instance is

$$f_{s_{k,n}}(s; \boldsymbol{\theta}_s) = \sum_{i=1}^{n_k} \phi_{k,i} \mathcal{N}(s; 0, \sigma_{k,i}^2) \quad \forall n = 1, \dots, N .$$
 (2)

The notation $\mathcal{N}(\boldsymbol{\xi}; \boldsymbol{\mu}, \boldsymbol{\Lambda})$ represents a normal distribution function with variable $\boldsymbol{\xi}$, mean $\boldsymbol{\mu}$, and covariance matrix $\boldsymbol{\Lambda}$. The number of Gaussians for the *k*th source is denoted by n_k . The variance and the weighting coefficients of the *i*th Gaussian are denoted by $\sigma_{k,i}^2$ and $\phi_{k,i}$, respectively. The weighting coefficients, $\phi_{k,i}$, satisfy $\sum_{i=1}^{n_k} \phi_{k,i} = 1, 0 < \phi_{k,i} \le 1, \forall k = 1, \dots, K$. For simplicity of the derivations, we assume that the means of the Gaussians are equal to zero. Extension to the nonzero case is straight-forward. Under the assumption of independent source signals, the joint PDF of the sources is modeled by a multivariate GMM with diagonal covariance matrices:

$$f_{\mathbf{s}_n}(\mathbf{s}_n;\boldsymbol{\theta}_{\mathbf{s}}) = \prod_{k=1}^{K} f_{s_k}(s_k;\boldsymbol{\theta}_{s_k}) = \sum_{m=1}^{M} \pi_m \mathcal{N}(\mathbf{s};\mathbf{0},\mathbf{C}_m) \quad (3)$$

where $M = \prod_{k=1}^{K} n_k$ is the GMM order, assumed to be known. The index *m* denotes a single combination of Gaussians from all the sources. The weighting coefficients, $\{\pi_m\}_{m=1}^{M}$, also satisfy $0 < \pi_m \leq 1, \sum_{m=1}^{M} \pi_m = 1 \ \forall m = 1, \dots, M$. The diagonal matrix \mathbf{C}_m represents the variances of the *m*th Gaussian. The set of unknown parameters is denoted by $\boldsymbol{\theta}_{\mathbf{s}} = \{\pi_m; \mathbf{C}_m\}_{m=1}^{M}$.

2.3. Sensors distribution model

Let $\mathbf{x} \stackrel{\triangle}{=} [\mathbf{x}_1^T, \dots, \mathbf{x}_N^T]^T$ denote the measurements at the *N* time instances. The sequence $\{\mathbf{x}_n\}_{n=1}^N$ is a *P*-order Markov process with known initial conditions, and thus, the PDF of \mathbf{x} can be expressed as $f_{\mathbf{x}}(\mathbf{x}) = \prod_{n=1}^N f_{\mathbf{x}_n | \mathbf{x}_n^P}(\mathbf{x}_n | \mathbf{x}_n^P)$. Since a linear transformation of a GMM-distributed random variable is also GMM-distributed, the conditional PDF of $\mathbf{x}_n | \mathbf{x}_n^P$ is also GMM, and using (1) and (3), it can be written in the form

$$f_{\mathbf{x}_n | \mathbf{x}_n^P}(\mathbf{x}_n | \mathbf{x}_n^P; \boldsymbol{\theta}_{\mathbf{x}}') = \sum_{m=1}^M \pi_m \mathcal{N}(\mathbf{x}_n; \mathbf{A}\mathbf{x}_n^P, \mathbf{R}_m)$$
(4)

where $\theta'_{\mathbf{x}} = \{\{\pi_m, \mathbf{C}_m\}_{m=1}^M, \mathbf{A}, \mathbf{H}\}\$ denotes the set of unknown distribution parameters of the observation signals and the structured covariance matrices are

$$\mathbf{R}_m \stackrel{\triangle}{=} \mathbf{H} \mathbf{C}_m \mathbf{H}^T, \ m = 1, \dots, M .$$
 (5)

In the case of lower number of sources than sensors, the matrix \mathbf{R}_m is singular because $rank(\mathbf{H}) \leq K < L$. In this case, \mathbf{R}_m can be replaced with $\mathbf{R}_m = \lim_{\varepsilon \to 0} (\mathbf{H}\mathbf{C}_m\mathbf{H}^T + \varepsilon \mathbf{I}_L)$ where \mathbf{I}_L is an identity matrix of size L.

2.4. Objective

In this paper, we are interested in two problems: 1) system identification, i.e. estimation of the state transition matrix **A** and input mixing matrix **H**, and 2) source separation, i.e. estimation of the source signals $\{s_n\}_{n=1}^N$. Estimation of the system parameters is performed in two stages. In the first stage, the state transition matrix, **A**, is estimated via the GEM algorithm. In the second stage, the input mixing matrix, **H**, is estimated and the signals are separated using the GMMJD algorithm [10]. In the SISO case the solution of the problem stated above is performed in similar stages: 1) estimation of the AR coefficients using the Yule Walker equations, and 2) estimation of the input signal variance. These stages are described in the following two sections.

3. STATE TRANSITION MATRIX ESTIMATION

In this section, the covariance matrices $\{\mathbf{R}_m\}_{m=1}^M$ are assumed to be unstructured, that is, the structure in (5) is ignored. Then, the log-likelihood function for estimation of $\boldsymbol{\theta}_{\mathbf{x}} = \{\{\pi_m, \mathbf{R}_m\}_{m=1}^M, \mathbf{A}\}$ from the observation vector \mathbf{x} , defined by the model presented in (1), can be written as:

$$\log f_{\mathbf{x}}(\mathbf{x}; \boldsymbol{\theta}_{\mathbf{x}}) = \sum_{n=1}^{N} \log \sum_{m=1}^{M} \pi_m \mathcal{N}(\mathbf{x}_n; \mathbf{A}\mathbf{x}_n^P, \mathbf{R}_m) .$$
(6)

Since the maximization of the above log-likelihood function w.r.t. the unknown parameters cannot be analytically performed, the GEM algorithm is used. In this GEM algorithm, the maximization step is replaced with the method of coordinate ascent which converges to a local maximum of the function [11].

The GEM algorithm can be implemented in problems where there is "complete" data $\mathcal{Z} = (\mathbf{x}, \mathbf{y})$ such that the expectation $E\{\log f_{\mathcal{Z}}(\mathcal{Z}; \boldsymbol{\theta}') | \mathbf{y}; \boldsymbol{\theta}''\}$ can be easily computed for any two parameters sets $\boldsymbol{\theta}', \boldsymbol{\theta}''$. In the GEM approach $\hat{\boldsymbol{\theta}}^{(i+1)}$ is computed such that $U\left(\hat{\boldsymbol{\theta}}^{(i+1)}, \hat{\boldsymbol{\theta}}^{(i)}\right) \geq U\left(\hat{\boldsymbol{\theta}}^{(i)}, \hat{\boldsymbol{\theta}}^{(i)}\right)$ where

$$U\left(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^{(i)}\right) \stackrel{\text{def}}{=} E_{\mathbf{y}|\mathbf{x}} \left[\log f_{\mathbf{x}, \mathbf{y}}(\mathbf{x}, \mathbf{y}; \boldsymbol{\theta}) | \mathbf{x}; \hat{\boldsymbol{\theta}}^{(i)} \right]$$
(7)

and i denotes the iteration index. Under fairly general conditions, this algorithm is guaranteed to converge to (at least) a local maximum of the log-likelihood function [12]. Therefore, it is referred to as QML estimator.

In our problem, the complete data is chosen to be $\mathcal{Z} = (\mathbf{x}, \mathbf{y})$ with $\mathbf{y} = [\mathbf{y}_1^T, \dots, \mathbf{y}_N^T]^T$, where the hidden indication vectors, $\mathbf{y}_n = (y_{n,1}, \dots, y_{n,m})^T$ $n = 1, \dots, N$ are randomized according to the following PDF:

$$f_{\mathbf{y}_n}(\mathbf{y}_n) = \sum_{m=1}^M \pi_m \delta(y_{n,m} - 1)$$
(8)

where δ denotes the Dirac's delta function and

$$y_{n,m} = \begin{cases} 1 & \text{if } \mathbf{s}_n \text{ is generated by the } m \text{th Gaussian} \\ 0 & \text{otherwise} \end{cases}$$
(9)

According to the Bayes theorem [13]:

$$f_{\mathbf{x}_n | \mathbf{x}_n^P}(\mathbf{x}_n | \mathbf{x}_n^P; \boldsymbol{\theta}_{\mathbf{x}}) = E_{\mathbf{y}}[f_{\mathbf{x}_n | \mathbf{y}_n, \mathbf{x}_n^P}(\mathbf{x}_n | \mathbf{y}_n, \mathbf{x}_n^P; \boldsymbol{\theta}_{\mathbf{x}})]$$
$$= \sum_{m=1}^M \pi_m f_{\mathbf{x}_n | \mathbf{y}_n, \mathbf{x}_n^P}(\mathbf{x}_n | \mathbf{y}_n = 1, \mathbf{x}_n^P; \boldsymbol{\theta}_{\mathbf{x}})$$

It can easily be seen that (10) is identical to (4). Using (8)-(10), the logarithm of the joint PDF of \mathbf{x}, \mathbf{y} is given by [13]:

$$\log f_{\mathbf{x},\mathbf{y}}(\mathbf{x},\mathbf{y};\boldsymbol{\theta}_{\mathbf{x}}) = \sum_{n=1}^{N} \sum_{m=1}^{M} y_{n,m} \log[\pi_{m} \mathcal{N}(\mathbf{x}_{n};\mathbf{A}\mathbf{x}_{n}^{P},\mathbf{R}_{m})].$$
(10)

According to (7), the expectation step in the GEM algorithm is the computation of the conditional expectation of (10):

$$U\left(\boldsymbol{\theta}_{\mathbf{x}}, \hat{\boldsymbol{\theta}}_{\mathbf{x}}^{(i)}\right) = \sum_{n=1}^{N} \sum_{m=1}^{M} \gamma_{n,m}^{(i)} \log[\pi_{m} \mathcal{N}(\mathbf{x}_{n}; \mathbf{A}\mathbf{x}_{n}^{P}, \mathbf{R}_{m})] \quad (11)$$

where $\gamma_{n,m}^{(i)} \triangleq E_{\mathbf{y}|\mathbf{x}} \left[y_{n,m} | \hat{\boldsymbol{\theta}}_{\mathbf{x}}^{(i)} \right]$. Since $y_{n,m}$ can have only discrete values of 0 and 1, by applying the Bayes theorem, $\gamma_{n,m}^{(i)}$ can be calculated in the following manner [13]:

$$\gamma_{n,m}^{(i)} = \frac{\hat{\pi}_m^{(i)} \mathcal{N}(\mathbf{x}_n; \hat{\mathbf{A}}^{(i)} \mathbf{x}_n^P, \hat{\mathbf{R}}_m^{(i)})}{\sum_{m=1}^M \hat{\pi}_m^{(i)} \mathcal{N}(\mathbf{x}_n; \hat{\mathbf{A}}^{(i)} \mathbf{x}_n^P, \hat{\mathbf{R}}_m^{(i)})} .$$
(12)

The maximization step in the GEM algorithm is performed by maximizing (11) w.r.t. $\theta_{\mathbf{x}}$. The mixture weights, π_m , are estimated by maximizing (11) under the constraint $\sum_{m=1}^{M} \pi_m = 1$:

$$\hat{\pi}_m^{(i+1)} = \frac{1}{N} \sum_{n=1}^N \gamma_{n,m}^{(i)} \quad \forall m = 1, \dots, M .$$
(13)

The covariance matrices, $\{\mathbf{R}_m\}_{m=1}^M$, are estimated by maximizing (11) w.r.t. $\boldsymbol{\theta}_{\mathbf{x}}$:

$$\hat{\mathbf{R}}_{m}^{(i+1)} = \frac{\sum_{n=1}^{N} \gamma_{n,m}^{(i)} (\mathbf{x}_{n} - \hat{\mathbf{A}}^{(i)} \mathbf{x}_{n}^{P}) (\mathbf{x}_{n} - \hat{\mathbf{A}}^{(i)} \mathbf{x}_{n}^{P})^{T}}{\sum_{n=1}^{N} \gamma_{n,m}^{(i)}} \quad (14)$$

 $\forall m = 1, \dots, M$. Maximization of (11) with respect to **A** is achieved via equating the corresponding partial derivatives to zero, which yields:

$$\sum_{n=1}^{N} \sum_{m=1}^{M} \gamma_{n,m}^{(i)} \hat{\mathbf{R}}_{m}^{(i)^{-1}} \mathbf{x}_{n} (\mathbf{x}_{n}^{P})^{T} = \sum_{n=1}^{N} \sum_{m=1}^{M} \gamma_{n,m}^{(i)} \hat{\mathbf{R}}_{m}^{(i)^{-1}} \hat{\mathbf{A}}^{(i+1)} \mathbf{x}_{n}^{P} (\mathbf{x}_{n}^{P})^{T}$$
(15)

By applying the *vec* operator on both sides of (15) and using the Kronecker products property $vec(\mathbf{A}_1\mathbf{A}_2\mathbf{A}_3) = (\mathbf{A}_3^T \otimes \mathbf{A}_1)vec(\mathbf{A}_2)$, we obtain:

$$vec(\hat{\mathbf{A}}^{(i+1)}) = \left[\sum_{m=1}^{M} \mathbf{T}_{m}^{(i)} \otimes \hat{\mathbf{R}}_{m}^{(i)^{-1}}\right]^{-1} vec\left[\sum_{m=1}^{M} \hat{\mathbf{R}}_{m}^{(i)^{-1}} \mathbf{G}_{m}^{(i)}\right]$$
(16)
where $\mathbf{T}_{m}^{(i)} \triangleq \sum_{n=1}^{N} \gamma_{n,m}^{(i)} \mathbf{x}_{n}^{P} (\mathbf{x}_{n}^{P})^{T}$ and $\mathbf{G}_{m}^{(i)} \triangleq \sum_{n=1}^{N} \gamma_{n,m}^{(i)} \mathbf{x}_{n} (\mathbf{x}_{n}^{P})^{T}$.
Note that due to the summation over the Gaussians the matrix

Note that due to the summation over the Gaussians, the matrix $\sum_{m=1}^{M} \mathbf{T}_{m}^{(i)} \otimes \hat{\mathbf{R}}_{m}^{(i)^{-1}}$ is not necessarily singular if the number of sensors is lower than the number of sources.

The GEM algorithm runs iteratively by carrying out the following two steps at each iteration:

- The E-step (Expectation): Compute $\gamma_{n,m}$ by (12).
- The M-step (Maximization): Maximizing (11) w.r.t. $\theta_{\mathbf{x}}$ with $\gamma_{n,m}$ obtained in the E-step.

These two steps are repeated till a predefined convergence criterion is satisfied.

It can be easily seen that in the case of MIMO-AR system with Gaussian distributed sources, this solution reduces to the well known multidimensional Yule-Walker equations obtained in the linear predictive coding (LPC) [14]:

$$\hat{\mathbf{A}}_{LPC} = \left[\sum_{n=1}^{N} \mathbf{x}_n (\mathbf{x}_n^P)^T\right] \left[\sum_{n=1}^{N} \mathbf{x}_n^P (\mathbf{x}_n^P)^T\right]^{-1} .$$
 (17)

The above procedure also coincides with the SISO-AR system identification technique, described in [9].

4. INPUT MIXING MATRIX ESTIMATION AND SOURCE SEPARATION

The proposed method is based on the coordinate ascent technique, where each iteration consists of two stages: 1) freeze \mathbf{H} and $\{\mathbf{C}_m\}_{m=1}^M$ to construct $\{\mathbf{R}_m\}_{m=1}^M$ and estimate $\{\pi_m\}_{m=1}^M$ and \mathbf{A} ; 2) freeze \mathbf{A} and $\{\pi_m\}_{m=1}^M$ to estimate \mathbf{H} and $\{\mathbf{C}_m\}_{m=1}^M$. The first stage can be performed according to the procedure described in the previous section in which estimation of $\{\mathbf{R}_m\}_{m=1}^M$ is replaced by (5). In other words, the extended algorithm is identical to the GEM algorithm presented in the previous section, where the maximization w.r.t. $\{\mathbf{R}_m\}_{m=1}^M$ is replaced by maximization w.r.t. \mathbf{H} and $\{\mathbf{C}_m\}_{m=1}^M$.

In this section, we focus on the second stage which provides an estimate of **H** and $\{\mathbf{C}_m\}_{m=1}^M$. The estimation in this part utilizes

the structure of $\{\mathbf{R}_m\}_{m=1}^M$ stated in (5) by implementing a joint diagonalization technique. Estimation of the matrix \mathbf{H} can be obtained by maximizing (6) w.r.t. \mathbf{H} and $\{\mathbf{C}_m\}_{m=1}^M$, in which the estimate of \mathbf{A} from the previous iteration is substituted.

By denoting $\mathbf{z}_n = \mathbf{x}_n - \mathbf{A}\mathbf{x}_n^P = \mathbf{H}\mathbf{s}_n, \ n = 1, \dots, N$, (6) can be rewritten as:

$$\log f_{\mathbf{x}}(\mathbf{x}; \boldsymbol{\theta}_{\mathbf{x}}) = \sum_{n=1}^{N} \log \sum_{m=1}^{M} \pi_m \mathcal{N}(\mathbf{z}_n; \mathbf{0}, \mathbf{R}_m) .$$
(18)

Maximization of (18) is identical to the maximization of the loglikelihood in the instantaneous BSS problem. In the BSS problem, the sources vector, s_n , is reconstructed by estimating a separation matrix, **B**, for which

$$\hat{\mathbf{s}}_n = \mathbf{B}\mathbf{z}_n, \quad \forall n = 1, \dots, N$$
 (19)

In [10] an ML based approach for estimating **B** was derived. This estimator is given by following minimization problem:

$$\mathbf{B} = \arg\min_{\mathbf{B}} \left(\sum_{m=1}^{M} \hat{\pi}_m \left[\log |DIAG(\mathbf{B}\hat{\mathbf{R}}_m \mathbf{B}^T)| - \log |\mathbf{B}\hat{\mathbf{R}}_m \mathbf{B}^T| \right] \right)$$

where $|\cdot|$ denotes the determinant operator and $DIAG(\mathbf{B}\hat{\mathbf{R}}_m\mathbf{B}^T)$ denotes a diagonal matrix with the same diagonal elements of $\mathbf{B}\hat{\mathbf{R}}_m\mathbf{B}^T$.

The minimum of (20) is attained for a matrix **B** which jointly diagonalizes the estimated GMM covariance matrices. An approximate joint diagonalization algorithm, offered by Pham [1], which minimizes (20) w.r.t. **B** is applied in order to estimate the separation matrix. In the case of lower number of sources than sensors (K < L), a dimension reduction stage is required in order to use the above source separation procedure [10].

In summary, the proposed coordinate ascent algorithm for solving the BSS problem runs iteratively by carrying out the following two-steps at each iteration:

- 1. Estimate the process \mathbf{z}_n by substituting $\hat{\mathbf{A}}$ from the previous iteration: $\hat{\mathbf{z}}_n = \mathbf{x} \hat{\mathbf{A}} \mathbf{x}_n^P$; Estimate $\{\mathbf{C}_m\}_{m=1}^M$ and \mathbf{B} and separate the source signals by applying the GMMJD method [10] that includes a joint diagonalization algorithm, offered by Pham [1].
- 2. Reconstruct $\{\mathbf{R}_m\}_{m=1}^M$ according to (5) and estimate π_m and **A** according to (12), (13) and (16).

These two steps are repeated till a predefined convergence criterion is satisfied. The matrix $\hat{\mathbf{A}}^{(0)}$ was computed using the LPC technique (17).

5. SIMULATIONS RESULTS

The performances of the proposed system identification and BSS techniques are evaluated via simulations with synthetic and speech data. Estimation performance of the transition matrix \mathbf{A} is compared to the multi-dimensional LPC algorithm presented in (17) which is the ML estimator under Gaussian assumption, and the BSS performance is compared to the FastICA [15] applied to the whitened signal where the whitening procedure is based on the estimated \mathbf{A} via LPC.

The estimation performance of **A** is evaluated via the normalized MSE, defined as $\hat{\mathbf{A}}_{MSE} = \frac{||\tilde{\mathbf{A}} - \hat{\mathbf{A}}||_F}{||\tilde{\mathbf{A}}||_F}$, where $|| \cdot ||_F$ denotes the Frobenius norm and $\tilde{\mathbf{A}} = [\mathbf{I} \quad \mathbf{A}]$ is the extended transition matrix which includes also the zeroth order coefficients matrix. The performance of the BSS algorithm is evaluated in terms of signal-tointerference ratio (SIR), defined as the ratio between the target signal power to the interference signal power: $SIR = \frac{|\hat{\mathbf{s}}^H \mathbf{s}|^2}{2}$

power to the interference signal power: $SIR = \frac{\|\hat{\mathbf{s}}^H \mathbf{s}\|^2}{\|\|\hat{\mathbf{s}}\|^2 - \|\|\hat{\mathbf{s}}^H \mathbf{s}\|^2}$. In the simulations, the case of K = 2 sources and L = 4 sensors was considered. The source signals were synthesized and mixed by

was considered. The source signals were synthesized and mixed by $\mathbf{C} = \begin{bmatrix} 0.57 & 0.51 & 0.43 & 0.49 \\ 0.51 & 0.52 & 0.49 & 0.48 \end{bmatrix}^T$, and an AR process of order P = 3 was generated by the state transition matrix

$$\mathbf{A} = 10^{-2} \begin{bmatrix} 16 & -7 & 4 & 5.5 & 8 & -8 \\ 10 & 4 & -7 & 6 & 7 & -13 \\ 15 & -12 & -1.5 & 1 & -2.5 & 15 \\ 6 & 6 & 4 & 5 & 2 & 15 \\ 9 & 2 & 5 & 1 & -2 & 1 \\ 14 & 6 & 5 & 3 & -2 & 1 \\ 12 & 10.5 & 8 & 12.5 & 11 & 2.5 \\ 6 & -6 & 17 & 4 & -4 & 2 \end{bmatrix}.$$

The estimation performances are evaluated using 500 Monte-Carlo trials. In the first example, the sources were zero-mean GMMdistributed with order M = 3 for both sources. The weighting coefficients and the variances of the GMM-distributions for both sources were set as follows: $\{\phi_{1,m}\}_{m=1}^3 = (0.2, 0.6, 0.2), \sigma_{1,m}^2 = \{3, 30, 300\}, \{\phi_{2,m}\}_{m=1}^3 = (0.55, 0.15, 0.3), \sigma_{2,m}^2 = \{10, 100, 1\}$. Fig. 2 shows that the proposed algorithm outperforms the multidimensional LPC, in the state transition matrix estimation and source separation.

In the second example, the performance of the proposed algorithm was evaluated using speech signals from the ICA'99 synthetic benchmarks [16] database, convolved with the MIM-AR system from the previous example, to obtain 4 mixed signals. The sampling frequency was $f_s = 22.05 kHz$. The observed mixed signals were segmented into frames with 50% overlapping. The separation algorithms were applied to each frame independently and the performances were evaluated based on the SIR averaged over the frames. A good qualitative recovery is confirmed by subjective listening to the recovered audio signals and by the average SIR performance presented in the following table:

Time [sec]	FastICA [dB]	MIMO-AR-GMM [dB]
0.25	1.84	6.51
0.5	3.36	8.92
2	5.23	11.31

6. CONCLUSION

A new, ML-based algorithm for BSS of convolutive mixtures modeled by MIMO-AR in the time domain was presented. The algorithm assumes GMM distribution of the sources. The state transition matrix and the sensors distribution parameters are jointly estimated by applying the GEM algorithm. The separation matrix and the sources are estimated via approximate joint diagonalization of the GMM covariance matrices. The proposed method extends the approach presented in [9] for SISO-AR system identification to MIMO-AR systems. The method was tested via simulations with synthetic and speech data. It was shown that the proposed method outperforms the multi-dimensional LPC in estimation of the state transition matrix, and results in good separation performances for convolutive mixtures.

7. REFERENCES

 D. Pham and J. Cardoso, "Blind separation of instantaneous mixtures of nonstationary sources," *IEEE Trans. Signal Processing*, vol. 49, no. 9, pp. 1837-1848, 2001.

- [2] D. Farina, C. Fevotte, C. Doncarli, and R. Merletti, "Blind separation of linear instantaneous mixtures of nonstationary surface myoelectric signals," *IEEE Trans. Biomedical Engineering*, vol. 51, no. 9, pp. 1555-1567, 2004.
- [3] N. Mitianoudis and M. Davies, "Audio source separation of convolutive mixtures," *IEEE Trans. Speech and Audio Processing*, vol. 11, no. 5, pp. 489-497, 2003.
- [4] H. Sawada, R. Mukai, S. Araki, and S. Makino, "A robust and precise method for solving the permutation problem of frequency-domain blind source separation," *IEEE trans. Speech and Audio Processing*, vol. 12, no. 5, pp. 530-538, 2004.
- [5] H. Buchner, R. Aichner, and W. Kellermann, "A generalization of a class of blind source separation algorithms for convolutive mixtures," *Proc. IEEE Int. Symposium on Independent Component Analysis and Blind Source Separation* (ICA), Nara, Japan, April 2003.
- [6] I. Russell, A. Mertins, and J. Xi, "Time domain optimization techniques for blind separation of non-stationary convolutive mixed signals," pp. 440-445, 2003.
- [7] Q. Li and A. Barron, "Mixture density estimation," Advances in Neural Information Processing Systems, vol. 12, pp. 279-285, MIT press, 2000.
- [8] K. E. Hild, H. T. Attias, and S.S. Nagarajan, "An EM method for spatiotemporal blind source separation using an ARMOG source model," *Proc. IEEE Int. Symposium on Independent Component Analysis and Blind Source Separation (ICA)*, 2006.
- [9] S. Verbout, J. Ooi, J. Ludwig, and A.V. Oppenheim, "Parameter estimation for autoregressive Gaussian-mixture processes :The EMAX algorithm," *IEEE Trans. Signal Processing*, vol. 46, no. 10, pp. 2744-2756, 1998.
- [10] K. Todros and J. Tabrikian, "Blind separation of independent sources using Gaussian mixture model," *IEEE Trans. Signal Processing*, in press.
- [11] D. G. Luenberger, *Linear and nonlinear programming*, 2nd ed. Reading, MA: Addison Wesley, 1984.
- [12] A. P. Dempster, N. M. Laird and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," J. R. Statist. Soc., Series B, pp. 1-38, 1977.
- [13] J.A. Bilmes, "A gentle tutorial of the EM algorithm and its application to parameter estimation for Gaussian mixture and hidden Markov models," *Technical Report TR-97-021*, International Computer Science Institute, Berkeley, California, 1998.
- [14] S. M. Kay, Modern spectral estimation: theory and application, Englewood Cliffs: PrenticeHall, 1988.
- [15] A. Hyvrinen, "Fast and robust fixed-point algorithms for independent component analysis," *IEEE Trans. on Neural Networks*, vol. 10, no. 3, pp. 626-634, 1999.
- [16] http://sound.media.mit.edu/ica-bench/