BLIND SEPARATION OF UNDERDETERMINED LINEAR MIXTURES BASED ON SOURCE NONSTATIONARITY AND AR(1) MODELING

Ondřej Šembera[1], *Petr Tichavský*[1] *and Zbyněk Koldovský*[2]

[1]Institute of Information Theory and Automation,
P.O.Box 18, 182 08 Prague 8, Czech Republic
[2]Technical University of Liberec, Czech Republic

ABSTRACT

The problem of blind separation of underdetermined instantaneous mixtures of independent signals is addressed through a method relying on nonstationarity of the original signals. The signals are assumed to be piecewise stationary with varying variances in different epochs. In comparison with previous works, in this paper it is assumed that the signals are not i.i.d. in each epoch, but obey a first-order autoregressive model. This model was shown to be more appropriate for blind separation of natural speech signals. A separation method is proposed that is nearly statistically efficient (approaching the corresponding Cramér-Rao lower bound), if the separated signals obey the assumed model. In the case of natural speech signals, the method is shown to have separation accuracy better than the state-of-the-art methods.

Index Terms— Autoregressive Processes, Cramér-Rao Bound, Blind Source Separation

1. INTRODUCTION

Blind separation of underdetermined instantaneous mixtures of independent signals is a challenging task - it is a nontrivial extension of independent component analysis (ICA) to the case when the number of sources exceeds the number of the available mixtures. In EEG signal processing, ICA was proven to be useful to separate parasitic signals (artifact) from the true physiological signals, and separate some physiological signals from the other [1]. However, it is unclear what the number of the components should be, and one can always expect that this number is greater than the number of the electrodes. In acoustics, the number of signals to be separated can also exceed that of microphones. Therefore the underdetermined mixtures are of great importance.

First methods to separate underdetermined mixtures were proposed by Comon and co-workers [2] [3] [4]. They are based on canonical polyadic (CP) decompositions of sample high-order cumulants of the mixed signals. Another important step in this direction was the SOBIUM algorithm (Second Order Blind Identification of Underdetermined Mixtures) [5], which performs the CP decomposition, through an approximate joint diagonalization (AJD) of lagged covariance matrices. The AJD is a well studied problem and several efficient algorithms exist [6]. Another algorithm for separating underdetermined mixtures is UDSEP (Under-Determined SEParation), which is sometimes called WASOBIUM (Weight-Adjusted SOBIUM)[7]. This algorithm is based on a modified CP decomposition of a tensor composed of covariance matrices of the partitioned mixture, using asymptotically optimum weights derived for the case that the signals are Gaussian i.i.d. in each epoch. The algorithm proposed in this paper can be seen as an extension of UDSEP.

For sake of completeness, the NAP-ACDC algorithm [8] should be mentioned, which is designed for blind separation of underdetermined mixtures. This algorithm is conceptually different than the previous two algorithms, because it relies on the assumption that the original signals are sparse in time. We include this algorithm in our simulation study.

The rest of the paper is organized as follows. Section 2 summarizes the signal model, Section 3 presents the proposed algorithm, Section 4 includes derivation of the Cramer-Rao lower bound for the separation problem, Section 5 contains simulation results and Section 6 concludes the paper.

2. SIGNAL MODEL

Consider an underdetermined instantaneous mixtures of quasi-stationary independent sources, i.e. the data follow an equation

$$\mathbf{x}_t = \mathbf{A}\mathbf{z}_t,\ t = 1 \dots N, \tag{1}$$

where $\mathbf{x}_t$ denotes the vector of the observed mixture at time t and $\mathbf{z}_t$ stands for the vector of the original signals at the given time. The matrix $\mathbf{A} \in \mathbb{R}^{d\times r}$ is called the mixing matrix. We will denote its elements by a_{ij}. In the underdetermined case, the number of mixtures d is smaller than the number of sources r. From now on, we assume that the mixing matrix has full row rank. Let $\mathbf{X} \in \mathbb{R}^{d\times N}$ and $\mathbf{Z} \in \mathbb{R}^{d\times N}$ be matrices whose columns are the time samples of the vector of mixtures and original signals, respectively. Then we can

This work was supported by the Czech Science Foundation through Project No. 14-13713S.

write the model equation in a matrix form

$$\mathbf{X} = \mathbf{A}\mathbf{Z}. \tag{2}$$

Our task is to estimate the mixing matrix and the source signals from the knowledge of the mixture matrix $\mathbf{X}$ only. We assume that each original signal $\{z_{j,t}\}$ can be partitioned into M blocks (indexed by m) of the same length T. In addition, we assume that in each block original signals are order one autoregressive processes with fixed variances and autoregressive coefficients, i.e.

$$z_{j,t} = q_{jm} z_{j,t-1} + \sigma_{jm} w_{j,t} \tag{3}$$

for

$$t = (m-1)T + 1, (m-1)T + 2, \dots mT.$$

Here q_{jm} denotes the autoregressive coefficient and σ_{jm} is a square root of the variance of the innovation sequence. The sequence $w_{j,t}$ is a zero mean and unit variance Gaussian white noise. We assume that different source signals are mutually independent and are independent in different blocks.

A covariance of the mixture within the m-th block is given by

$$\mathbf{R}_m = \mathbf{A}\,\mathrm{diag}(\mathbf{D}_{:,m})\mathbf{A}^\top, \tag{4}$$

where $\mathbf{A}^\top$ denotes a transpose of $\mathbf{A}$. The covariance with lag one $\mathbf{S}_m = \mathrm{E}\left[\mathbf{x}_t \mathbf{x}_{t-1}^\top\right]$ takes the form

$$\mathbf{S}_m = \mathbf{A}\,\mathrm{diag}(\mathbf{D}_{:,m})\,\mathrm{diag}(\mathbf{Q}_{:,m})\mathbf{A}^\top. \tag{5}$$

The matrices $\mathbf{D} \in \mathbb{R}^{r\times M}$ and $\mathbf{Q} \in \mathbb{R}^{r\times M}$ are composed of the sources variances and autoregressive coefficients, respectively. Note that the variance d_{jm} of the j-th source in the m-th block is different from the variance σ_{jm} of the innovation sequence and depends both on σ_{jm} and on q_{jm}. Throughout the paper, the vector of the parameters to-be estimated will be denoted by θ, which contains all elements of $\mathbf{A}$, $\mathbf{D}$ and $\mathbf{Q}$.

3. APPROXIMATE MLE

3.1. Weighted least squares criterion

In the following we describe a method of estimating the mixing matrix $\mathbf{A}$ using only the sample lag zero and lag one covariances, respectively, computed as

$$\hat{\mathbf{R}}_m = \frac{1}{T} \sum_{t=(m-1)T+1}^{mT} \mathbf{x}_t \mathbf{x}_t^\top, \tag{6}$$

$$\hat{\mathbf{S}}_m = \frac{1}{T-1} \sum_{t=(m-1)T+2}^{mT} \mathbf{x}_t \mathbf{x}_{t-1}^\top, \tag{7}$$

only. These estimates are statistically unbiased. Since the sample zero lag covariance matrices are symmetric, we can limit ourselves to use only the elements of their lower triangular parts without loosing any information about the estimated parameters. Similarly, we can use only the lower triangular parts of the symmetric part of the sample lag one covariances. By the symmetric parts we understand matrices $\mathbf{S}_m^s$ given by $\hat{\mathbf{S}}_m^s = 1/2(\hat{\mathbf{S}}_m + \hat{\mathbf{S}}_m^\top)$. [1] Let $\mathcal{L}$ denote a linear transformation vectorising the lower triangular part of a matrix argument, i.e.

$$\mathcal{L}(\mathbf{R}) = [r_{11}, r_{21} \dots r_{d1}, r_{22}, r_{32} \dots r_{d2}, r_{33} \dots]^\top \tag{8}$$

and let

$$\hat{\mathbf{v}}_m = \begin{bmatrix} \mathcal{L}(\hat{\mathbf{R}}_m) \\ \mathcal{L}(\hat{\mathbf{S}}_m^s) \end{bmatrix}. \tag{9}$$

Thanks to the central limit theorem, we can approximate the probability density of the statistics $\hat{\mathbf{v}}_m$ by a Gaussian distribution with a corresponding mean and covariance resulting in the density

$$p(\hat{\mathbf{v}}_1 \dots \hat{\mathbf{v}}_M) = \prod_{m=1}^{M} (2\pi)^{-d(d+1)/2} |\mathbf{C}_m|^{-1/2} \cdot \exp\left[-\frac{1}{2}\left(\hat{\mathbf{v}}_m - \mathbf{v}_m(\theta)\right)^\top \mathbf{C}_m^{-1} \left(\hat{\mathbf{v}}_m - \mathbf{v}_m(\theta)\right)\right], \tag{10}$$

where the mean $\mathbf{v}_m(\theta) = \mathrm{E}\left[\hat{\mathbf{v}}_m\right]$ takes the form

$$\mathbf{v}_m(\theta) = \begin{bmatrix} \mathcal{L}(\mathbf{A}\,\mathrm{diag}(\mathbf{D}_{:,m})\mathbf{A}^\top) \\ \mathcal{L}(\mathbf{A}\,\mathrm{diag}(\mathbf{D}_{:,m})\,\mathrm{diag}(\mathbf{Q}_{:,m})\mathbf{A}^\top) \end{bmatrix}. \tag{11}$$

$\mathbf{C}_m$ denotes the covariance matrix of $\hat{\mathbf{v}}_m$. $\mathbf{C}_m$ is composed of the crosscovariances of the elements of the sample covariance matrices, i.e. of the terms

$$\mathbf{Cov}\left[\hat{\mathbf{R}}_{ijm}, \hat{\mathbf{R}}_{uvm}\right] = \tfrac{1}{T} \sum_{\alpha,\beta=1}^{r} d_{\alpha m} d_{\beta m} (a_{i\alpha} a_{u\alpha} a_{j\beta} a_{v\beta} + a_{i\alpha} a_{v\alpha} a_{j\beta} a_{u\beta}) \cdot \mathbf{Q}_{\alpha\beta m}^{RR}, \tag{12a}$$

$$\mathbf{Cov}\left[\hat{\mathbf{R}}_{ijm}, \hat{\mathbf{S}}_{uvm}\right] = \tfrac{1}{T} \sum_{\alpha,\beta=1}^{r} d_{\alpha m} d_{\beta m} (a_{i\alpha} a_{u\alpha} a_{j\beta} a_{v\beta} + a_{i\alpha} a_{v\alpha} a_{j\beta} a_{u\beta}) \cdot \mathbf{Q}_{\alpha\beta m}^{RS}, \tag{12b}$$

$$\mathbf{Cov}\left[\hat{\mathbf{S}}_{ijm}, \hat{\mathbf{S}}_{uvm}\right] = \tfrac{1}{T} \sum_{\alpha,\beta=1}^{r} d_{\alpha m} d_{\beta m} a_{i\alpha} a_{u\alpha} a_{j\beta} a_{v\beta} \cdot \mathbf{Q}_{\alpha\beta m}^{RR} + d_{\alpha m} d_{\beta m} a_{i\alpha} a_{v\alpha} a_{j\beta} a_{u\beta} \cdot \mathbf{Q}_{\alpha\beta m}^{SS}, \tag{12c}$$

[1] The sample lag one covariance matrices can be written as a sum of their symmetric and asymmetric parts, $\hat{\mathbf{S}}_m = \hat{\mathbf{S}}_m^a + \hat{\mathbf{S}}_m^s$, where $\hat{\mathbf{S}}_m^s = 1/2(\hat{\mathbf{S}}_m + \hat{\mathbf{S}}_m^\top)$ and $\hat{\mathbf{S}}_m^a = 1/2(\hat{\mathbf{S}}_m - \hat{\mathbf{S}}_m^\top)$. Note that $\mathrm{E}\left[\hat{\mathbf{S}}_m^a\right] = 0$. Therefore $\hat{\mathbf{S}}_m^a$ approaches zero matrix as the length parameter T goes to infinity and the information content of the asymmetric part diminishes. As a result, the information about estimated parameters is contained predominantly in the lower triangular parts of the symmetric matrix $\hat{\mathbf{S}}_m^s$.

where $\hat{\mathbf{R}}_{ijm}$ denotes the i,j-th element of the zero lag sample covariance matrix $\hat{\mathbf{R}}_m$; the same holds for $\hat{\mathbf{S}}_{ijm}$. The auxiliary terms $\mathbf{Q}^{RR}_{\alpha\beta m}$, $\mathbf{Q}^{RS}_{\alpha\beta m}$ and $\mathbf{Q}^{SS}_{\alpha\beta m}$ are given by

$$\mathbf{Q}^{RR}_{\alpha\beta m} = \frac{1 + q_{\alpha m} q_{\beta m}}{1 - q_{\alpha m} q_{\beta m}}, \tag{13a}$$

$$\mathbf{Q}^{RS}_{\alpha\beta m} = \frac{q_{\alpha m} + q_{\beta m}}{1 - q_{\alpha m} q_{\beta m}}, \tag{13b}$$

$$\mathbf{Q}^{SS}_{\alpha\beta m} = q_{\alpha m} q_{\beta m} + \frac{q^2_{\alpha m} + q^2_{\beta m}}{1 - q_{\alpha m} q_{\beta m}}. \tag{13c}$$

They were derived by neglecting the terms of order smaller than $\mathcal{O}(T^{-1})$.

$\mathbf{A}$, $\mathbf{D}$ and $\mathbf{Q}$ are estimated by minimizing the negative of the log-likelihood function corresponding to the PDF (10). This is done using Levenberg-Marquardt iteration method. To simplify the computations, we exclude the covariance matrices $\mathbf{C}_m$ from the minimization and only re-evaluate them in every tenth iteration. In this way we obtain a weighted least squares criterion

$$\begin{aligned} Q_1(\theta) \quad = \quad & \tfrac{1}{2} \sum_{m=1}^{M} \left(\hat{\mathbf{v}}_m - \mathbf{v}_m(\theta)\right)^\top \mathbf{W}_m \\ & \left(\hat{\mathbf{v}}_m - \mathbf{v}_m(\theta)\right). \end{aligned} \tag{14}$$

where $\mathbf{W}_m = (\mathbf{C}_m + \varepsilon \mathbf{I})^{-1}$. The scalar parameter ε is gradually decreased to zero as the algorithm proceeds.

3.2. Nonnegativity and stationarity constraints

To prevent the optimization method stepping out of the feasible area given by inequalities $d_{jm} > 0$, $-1 < q_{jm} < 1$ for any $j = 1 \dots r$ and $m = 1 \dots M$, we add a barrier function to the criterion resulting in the final cost function

$$\begin{aligned} Q_2(\theta) = Q_1(\theta) - \alpha \sum_{j=1}^{r} \sum_{m=1}^{M} \big[& \log d_{jm} \\ & + \log(1 + q_{jm}) + \log(1 - q_{jm}) \big]. \end{aligned} \tag{15}$$

The scalar parameter α is gradually decreased to zero.

3.3. Optimization details

The algorithm is initialized by the output of the UDSEP algorithm [7]. Then it proceeds with iterations of the Levenberg-Marquardt method [9].

In short, the update step of the method is $\theta \leftarrow \theta - (\mathbf{H} + \mu \mathbf{I})^{-1} \mathbf{g}$, where $\mathbf{g}$ is the gradient of the criterion (15) and $\mathbf{H}$ is its approximate hessian. In every tenth iteration, we update the covariance matrices $\mathbf{C}_m$, barrier function parameter α and the regularization parameter ε. Pseudocode of the algorithm is summarized in Algorithm 1.

Algorithm 1 Optimization method

```
Initialize A, D, Q and parameters α, ε;
Compute weight matrices W_m = (C_m + εI)^{-1};
it = 0;
while it ≤ it_max do
    Perform the Levenberg-Marquardt update of A, D and Q;
    it = it + 1;
    if it mod 10 = 0 then
        α = α/2, ε = ε/2;
        Update covariance matrices C_m;
        Re-compute W_m = (C_m + εI)^{-1};
    end if
end while
```

4. ACCURACY OF THE ESTIMATE AND ITS CRAMÉR-RAO LOWER BOUND

The accuracy of the estimate $\hat{\mathbf{A}}$ is assessed by the sum of square angular errors of the estimates of its columns. Let ω_j denote the angle between the j-th column of the true mixing matrix and its estimate, i.e.

$$\cos \omega_j = \frac{\mathbf{A}^\top_{:,j} \hat{\mathbf{A}}_{:,j}}{\|\mathbf{A}_{:,j}\| \|\hat{\mathbf{A}}_{:,j}\|}. \tag{16}$$

Before computing (16), the columns of $\hat{A}$ must be reordered to match the corresponding columns of the true mixing matrix. Then, we measure the accuracy by the sum $MSAE = \sum_{j=1}^{r} \omega_j^2$.

The Cramér-Rao lower bound for $MSAE$ was derived in [7]. Its computation requires the knowledge of the Fisher information matrix for the unknown parameter (mixing matrix $\mathbf{A}$). There are three different Fisher information matrices to be considered corresponding to different statistics used to estimate model parameters. This topic was discussed in [10]. First, the density of the whole data sample $\mathbf{X}$ has the form

$$\begin{aligned} p(\mathbf{X}) \quad = \quad & \prod_{m=1}^{M} (2\pi)^{-dT/2} |\tilde{\mathbf{C}}_m|^{-1/2} \\ & \cdot \exp\left[-\tfrac{1}{2} {\mathbf{x}^{(m)}}^\top \tilde{\mathbf{C}}_m^{-1} \mathbf{x}^{(m)} \right]. \end{aligned} \tag{17}$$

Here $\mathbf{x}^{(m)}$ denotes vectorised m-th block of the mixtures and $\tilde{\mathbf{C}}_m$ is its covariance matrix given by the formula

$$\begin{aligned} \tilde{\mathbf{C}}_m \quad = \quad & \mathrm{bstoeplitz}(\mathbf{A} \operatorname{diag}(\mathbf{D}_{:,m}) \mathbf{A}^\top, \\ & \mathbf{A} \operatorname{diag}(\mathbf{D}_{:,m}) \operatorname{diag}(\mathbf{Q}_{:,m}) \mathbf{A}^\top, \dots \\ & \mathbf{A} \operatorname{diag}(\mathbf{D}_{:,m}) \operatorname{diag}(\mathbf{Q}_{:,m})^{(T-1)} \mathbf{A}^\top). \end{aligned} \tag{18}$$

Second, we can use the approximate density of the statistics $\hat{\mathbf{v}}_m$ given by (10). The third option is using similar approximation for the lower triangular parts of the sample covariances with lag zero. Corresponding Fisher information matrices will be denoted by F_X, F_{RS}, F_R and Cramér-Rao

induced bounds (for $MSAE$) by $CRIB_X$, $CRIB_{RS}$ and $CRIB_R$ respectively. The estimation accuracy of the algorithms which use only the lag zero covariance matrices and those using both lag zero and one should roughly match $CRIB_R$ and $CRIB_{RS}$, respectively.

As all considered densities are Gaussian, we can use the formula for Fisher information matrix of a Gaussian distribution with a differentiable mean $\mathbf{e}(\theta)$ and covariance matrix $\mathbf{C}(\theta)$,

$$\mathbf{F}_{\theta_i\theta_j} = \frac{\partial \mathbf{e}}{\partial \theta_i}^\top \mathbf{C}^{-1} \frac{\partial \mathbf{e}}{\partial \theta_j} + \frac{1}{2} \operatorname{tr}\left[\mathbf{C}^{-1}\frac{\partial \mathbf{C}}{\partial \theta_i}\mathbf{C}^{-1}\frac{\partial \mathbf{C}}{\partial \theta_j}\right]. \quad (19)$$

Note that in the case of F_X, the first term in (19) is zero. In the other two cases, F_{RS} and F_R, the second term is neglectable assuming large T, and only the first term is used.

5. SIMULATIONS

5.1. Artificial data obeying the given model

In the first experiment, four artificial signals of the length $N = 6000$ samples partitioned into $M = 6$ blocks of $T = 1000$ samples were generated in twenty independent trials as AR(1) processes with block-wise constant variances and AR coefficients. The mixing matrix has columns $\mathbf{A}_{:,j} = [1 \ \cos\phi_j \ \sin\phi_j]^\top$, where $\phi = [0,\ \pi/4,\ \pi/2,\ 3\pi/4]$. The AR coefficients follow the equation $\mathbf{Q} = q \cdot \mathbf{Q}_0$, where $\mathbf{Q}_0$ is a fixed matrix and q is a magnitude parameter varying from zero to one. To demonstrate the effectiveness of the algorithm in the case of the AR coefficients close to ± 1, all the entries of $\mathbf{Q}_0$ are chosen relatively close to ± 1,

$$\mathbf{Q}_0 = \begin{bmatrix} 1 & 0.8 & 1 & 0.5 & 0.99 & 0.88 \\ 0.99 & -0.9 & 0.9 & -0.95 & 1 & -1 \\ 1 & 1 & -1 & -1 & 1 & 0.7 \\ 1 & 1 & 1 & -1 & -1 & -1 \end{bmatrix}.$$

The results are displayed in Fig. 1a. Accuracy of the proposed method (denoted as AR1SEP) increases with the magnitude of the AR coefficients, while the accuracy of the competing algorithms decreases in agreement with increase of the $CRIB_R$. Different $CRIB$s for the same setting are compared in Fig. 1b.

5.2. Speech data separation

In each of the one hundred trials, we select four speech utterances of the length 8.375 s sampled at 16kHz from a set of 16 speech samples (the same dataset as in [7]) and mix them using the same mixing matrix as in the previous experiment.The algorithm was for thirteen different choices of the time length of a block ranging from 30 to 150 ms. The results are shown in Figures 1c and 1d. The accuracy is measured by a sample mean and a sample median of the summed square angular errors. The proposed algorithm yields better estimation results than the other three algorithms used in comparison.

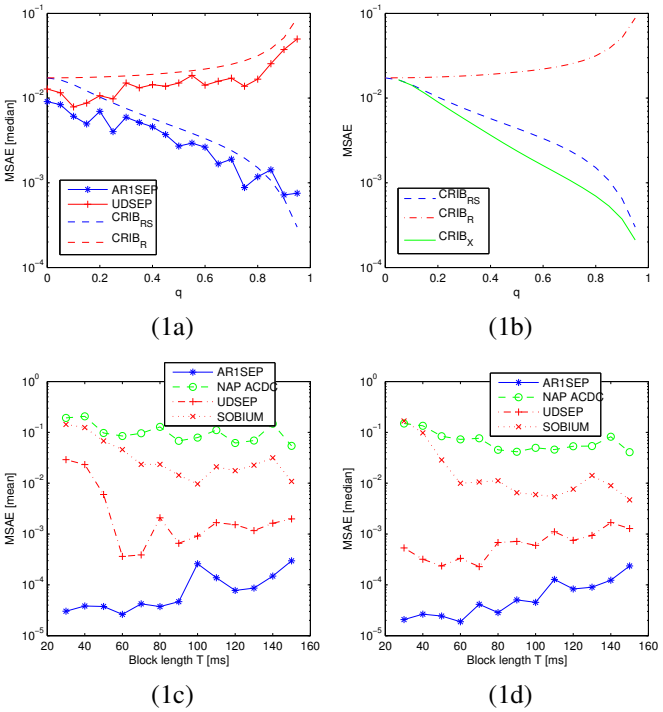

Fig. 1: (a) Application on artificial data following the given model. Dependency of $MSAE$ of the estimate on the AR coefficients magnitude parameter q and its comparison with $CRIB$. (b) Comparison of $CRIB$s corresponding to different densities used. (c,d) Speech data separation. Dependency of $MSAE$ of the estimate on the chosen length of the blocks.

6. CONCLUSION

We have proposed a novel algorithm for the blind separation of undetermined mixtures of nonstationary sources using lag zero and lag one covariances. The presented algorithm achieves better estimation of the mixing matrix for data following the assumed model when autoregressive parameters of the sources aproach one or minus one. The performance of the algorithms using only lag zero covariances decreases in this case, which is in accordance with the increase of the corresponding Cramér-Rao lower bound. In speech separation, our method achieves better estimation accuracy than the compared algorithms.

In future we plan to develop an on-line variant of the method. The code of the presented algorithm is available for download on the website of the second author.

7. REFERENCES

[1] M. Niknazar, H. Becker, B. Rivet, C. Jutten, P. Comon, "Blind source separation of underdetermined mixtures of event-related sources", *Signal Processing*, vol. 101, pp. 52-64, 2014

[2] P. Comon, "Blind Identification and Source Separation in 2x3 Under-Determined Mixtures," *IEEE Tr. Signal Processing*, vol. 52, no. 1, pp. 11–22, Jan. 2004.

[3] P. Comon and M. Rajih, "Blind Identification of Under-Determined Mixtures based on the Characteristic Function," *Signal Processing*, Elsevier, vol. 86, no. 9, pp. 2271-2281, 2006.

[4] M. Niknazar, H. Becker, B. Rivet, C. Jutten, and P. Comon, "Blind source separation of underdetermined mixtures of event-related sources," *Signal Processing*, vol. 101, pp. 52-64, August 2014.

[5] De Lathauwer, Lieven, and Josphine Castaing. "Blind identification of underdetermined mixtures by simultaneous matrix diagonalization." *Signal Processing, IEEE Transactions* on 56.3 (2008): 1096-1105.

[6] G. Chabriel, M. Kleinsteuber , E. Moreau, H. Shen, P. Tichavský, and A. Yeredor, "Joint Matrices Decompositions and Blind Source Separation", *IEEE Signal Processing Magazine*, vol. 31, pp. 34-43, 2014

[7] P. Tichavský, Z. Koldovský, "Weight adjusted tensor method for blind separation of underdetermined mixtures of nonstationary sources", *IEEE Tr. Signal Processing*, vol. 59, no. 3, pp. 1037–1047, March 2011.

[8] K.K. Lee, W.-K. Ma, X. Fu, T.-H. CHan, C.-Y. Chi, "A Khatri-Rao subspace approach to blind identification of mixtures of quasi-stationary sources", *Signal Processing*, vol. 93, no. 12, pp. 3515-3527, December 2013.

[9] K. Madsen, H. B. Nielsen, O. Tingleff, "Methods for non-linear least squares problems," *Lecture note, Informatics and Mathematical Modelling*, Technical University of Denmark, DTU, Lyngby. 2004.

[10] P. Tichavský, O. Šembera, and Z. Koldovský, "Blind separation of mixtures of piecewise AR(1) processes and model mismatch, " *Proc. Int. Conf. on Latent Variable Analysis and Signal Separation LVA/ICA 2015*, Liberec, Czech Republic, LNCS 9237, Springer 2015, pp. 304–311.

[11] P. Tichavský, A. Yeredor, and Z. Koldovský, "A Fast Asymptotically Efficient Algorithm for Blind Separation of a Linear Mixture of Block-Wise Stationary Autoregressive Processes," *ICASSP 2009*, pp. 3133-3136, Taipei, Taiwan, April 2009.