

A FAST ASYMPTOTICALLY EFFICIENT ALGORITHM FOR BLIND SEPARATION OF A LINEAR MIXTURE OF BLOCK-WISE STATIONARY AUTOREGRESSIVE PROCESSES

Petr Tichavský¹, Arie Yeredor² and Zbyněk Koldovský^{1,3}

¹Institute of Information Theory and Automation,
P.O.Box 18, 182 08 Prague 8, Czech Republic. E-mail: p.tichavsky@ieee.org

²School of Electrical Engineering, Tel-Aviv University
P.O.Box 39040, Tel-Aviv 69978, ISRAEL. E-mail: arie@eng.tau.ac.il

³Faculty of Mechatronics and Interdisciplinary Studies, Technical University of Liberec,
Studentská 2, 461 17 Liberec, Czech Republic. E-mail: zbynek.koldovsky@tul.cz

ABSTRACT

We propose a novel blind source separation algorithm called Block AutoRegressive Blind Identification (BARBI). The algorithm is asymptotically efficient in separation of instantaneous linear mixtures of blockwise stationary Gaussian autoregressive processes. A novel closed-form formula is derived for a Cramér Rao lower bound on elements of the corresponding Interference-to-Signal Ratio (ISR) matrix. This theoretical ISR matrix can serve as an estimate of the separation performance on the particular data. In simulations, the algorithm is shown to be applicable in blind separation of a linear mixture of speech signals.

Index Terms— Approximate joint diagonalization, blind source separation, autoregressive processes, second-order statistics

1. INTRODUCTION

Blind source separation aims to separate independent original signals from their instantaneous linear mixture, symbolically $\mathbf{X} = \mathbf{A}\mathbf{S}$, where \mathbf{S} is a matrix that contains, as rows, the original signals, \mathbf{A} is an unknown mixing matrix and \mathbf{X} represents the available data. Throughout this paper we assume that the mixing matrix has square form, say the dimension $d \times d$. In simulations, we also consider the overdetermined case, when the number of sources is smaller than the number of received signals.

Three possible routes to solution of the blind source separation can be found in the literature: non-Gaussianity, non-stationarity, and spectral diversity [3]. Many real-world signals exhibit all three of these features. This paper aims to

This work was supported by Ministry of Education, Youth and Sports of the Czech Republic through the project 1M0572 and by Grant Agency of the Czech Republic through the project 102/07/P384.

utilize jointly non-stationarity and spectral diversity like the papers [2, 4, 6, 7].

Unlike the previous algorithms, the proposed algorithm is asymptotically efficient, when the data obey the assumed piecewise stationary AR model, i.e. the resultant ISR matrix of the algorithm approaches the corresponding Cramer-Rao-induced lower bound, if the length of the blocks, where the sources are stationary, goes to infinity. In general, the algorithm is very fast compared to its competitors and allows to process data with high dimensions (100+). A novel expression for the CRB-induced bound is derived in Section 3. This expression allows to predict the accuracy of the blind separation for given set of the data. Moreover, the estimated ISR matrix allows to select the most “interesting” components to become the source estimates in the overdetermined case. In simulations in Section 4, the algorithm is tested on blind separation of instantaneous mixtures of artificial (model obeying) signals and natural speech signals, and its performance is compared to that of the algorithm by Pham [7].

2. THE PROPOSED ALGORITHM

Assume that the received signals can be divided into M blocks of the equal length (the extension of the method to blocks of unequal length is straightforward). In addition, assume that the blocks of the source signals can be modeled as Gaussian autoregressive random processes of the order that is smaller or equal to some maximum order p_{\max} . The blind source separator proposed in the paper will be obtained by a suitable approximate joint diagonalization (AJD) of the $M \cdot L$ matrices

$$\hat{\mathbf{R}}_{m\ell} = \frac{1}{2N_B} \{ \mathbf{X}^{(m,0)} [\mathbf{X}^{(m,\ell)}]^T + \mathbf{X}^{(m,\ell)} [\mathbf{X}^{(m,0)}]^T \}$$

$m = 1, \dots, M; \ell = 0, 1, \dots, p_{\max}$, where $L = p_{\max} + 1$, $N_B = N/M$ is the length of each block, N is the total length

of the data, and

$$\mathbf{X}^{(m,\ell)} = [\mathbf{X}_{:, (m-1)N_B + \ell + 1}, \dots, \mathbf{X}_{:, mN_B + \ell}]$$

is the m -th signal block of the data shifted to the right by ℓ samples. (For simplicity, we assume here that the data $\mathbf{X}(:, N+1), \dots, \mathbf{X}(:, N+p_{\max})$ were available.) The approximate joint diagonalization (AJD) means finding a *demixing* matrix $\widehat{\mathbf{V}}$ such that the matrices $\widehat{\mathbf{V}}\widehat{\mathbf{R}}_{m\ell}\widehat{\mathbf{V}}^T$ are all roughly diagonal. Most of existing algorithms are gradient-based and exhibit relatively slow convergence. Comparison of performance of these algorithms is even more complicated, because not all of them attempt to optimize the same criterion of the approximate diagonality.

The algorithm that is proposed in this paper is kind of approximate Newton algorithm, and therefore it exhibits nearly quadratic (very fast) convergence even in high dimensions. Moreover, each iteration is computationally cheap, having its complexity dominated by complexity of the transform $\widehat{\mathbf{R}}_{m\ell} \rightarrow \mathbf{V}\widehat{\mathbf{R}}_{m\ell}\mathbf{V}^T$ for all m and ℓ .

The algorithm is similar to the WASOBI algorithm (Weight-Adjusted SOBI) of [9, 11, 13] generalized to an arbitrary number of blocks, but is computationally simpler than the natural extension of WASOBI, called Block WASOBI [13], because it does not require computation of weight matrices used in WASOBI. Instead, it utilizes the Pham's-Garat's condition for optimum estimate[8], as it will be explained later.

Let $\widehat{\mathbf{r}}_{k\ell m}(\mathbf{V})$ denote the $L \times 1$ vector composed of (k, ℓ) -th elements of the matrices $\mathbf{V}\widehat{\mathbf{R}}_{mn}\mathbf{V}^T$, $n = 0, \dots, p_{\max}$, $m = 1, \dots, M$, and $k, \ell = 1, \dots, d$. Let $\{\mathbf{W}_{k\ell m}\}$ is a set of suitable positive definite weight matrices of the size $L \times L$. Note that the optimum weight matrices $\mathbf{W}_{k\ell m}$ for the WASOBI-like algorithm would be given as the inverse of $\text{cov}[\widehat{\mathbf{r}}_{k\ell m}(\mathbf{V})]$. In [11] these optimum weight matrices were defined in terms of the power spectra of the k -th and the ℓ -th source in the m -th interval. To be more specific, the paper [11] has dealt with separation of stationary signals as if $M = 1$, and the index m was omitted. It was shown in [13] that for all $k, \ell = 1, \dots, d$ and for the optimum weight matrices $\mathbf{W}_{k\ell}$ it holds that $\mathbf{W}_{k\ell}\mathbf{r}_{\ell\ell}(\mathbf{V}) = \mathbf{q}_k$ where \mathbf{q}_k only depends on the spectrum (AR coefficients) of the k -th source, and is independent of the spectrum of the ℓ -th source. In particular,

$$\mathbf{q}_k = \left[\frac{1}{2}\phi_k[0], \phi_k[1], \dots, \phi_k[p_{\max}] \right]^T$$

where $\phi_k[\tau]$ is an inverse of the Fourier transform of the inverse of the Fourier transform of the covariance function of the k -th source in the m -th interval, $\mathbf{r}_k[\tau] = \mathbf{r}_{k\tau}$. In other words, it holds that the convolution of $\mathbf{r}_k[\tau]$ and $\phi_k[\tau]$ is the Kronecker's delta. Here, however, one cannot restrict τ to the interval $[0, p_{\max}]$ but has to consider covariance extension of the AR process for $-\infty < \tau < \infty$. Therefore the proper computation of $\mathbf{q}_k^{[i]}$ involves estimation of autoregressive coefficients of the k -th source given the available covariances of the process.

The main iteration of WASOBI (where $M = 1$, index m is missing) is

$$\mathbf{V}^{[i+1]} = [\mathbf{A}^{[i]}]^{-1}\mathbf{V}^{[i]}$$

where i is the iteration index, the off-diagonal elements of $\mathbf{A}^{[i]}$ obey the 2×2 linear systems

$$\begin{bmatrix} \widehat{A}_{k\ell}^{[i]} \\ \widehat{A}_{\ell k}^{[i]} \end{bmatrix} = \begin{bmatrix} \widehat{\mathbf{r}}_{\ell\ell}^T \mathbf{W}_{k\ell} \widehat{\mathbf{r}}_{\ell\ell} & \widehat{\mathbf{r}}_{k\ell}^T \mathbf{W}_{k\ell} \widehat{\mathbf{r}}_{\ell\ell} \\ \widehat{\mathbf{r}}_{k\ell}^T \mathbf{W}_{k\ell} \widehat{\mathbf{r}}_{\ell\ell} & \widehat{\mathbf{r}}_{k\ell}^T \mathbf{W}_{k\ell} \widehat{\mathbf{r}}_{k\ell} \end{bmatrix}^{-1} \begin{bmatrix} \widehat{\mathbf{r}}_{\ell\ell}^T \mathbf{W}_{k\ell} \widehat{\mathbf{r}}_{k\ell} \\ \widehat{\mathbf{r}}_{k\ell}^T \mathbf{W}_{k\ell} \widehat{\mathbf{r}}_{k\ell} \end{bmatrix}, \quad (1)$$

and the diagonal elements of $\mathbf{A}^{[i]}$ are set to one. Note that the stationary point of the algorithm where the convergence is stopped, obeys relations $\widehat{\mathbf{r}}_{\ell\ell}^T \mathbf{W}_{k\ell} \widehat{\mathbf{r}}_{k\ell} = \widehat{\mathbf{r}}_{k\ell}^T \mathbf{W}_{k\ell} \widehat{\mathbf{r}}_{\ell\ell} = 0$ for all k, ℓ , $k \neq \ell$. These conditions are in accord with the condition $\mathbf{q}_k^T \widehat{\mathbf{r}}_{k\ell} = \mathbf{q}_\ell^T \widehat{\mathbf{r}}_{k\ell} = 0$, derived by Pham and Garat [8] for optimal separation of sources with known spectra, see [13].

In our algorithm, we replace the products $\mathbf{W}_{k\ell}\widehat{\mathbf{r}}_{\ell\ell}$ by \mathbf{q}_k and consider a general number M of intervals together to determine $\mathbf{A}^{[i]}$ (because the mixing matrix is assumed to be common to all intervals). The result is

$$\begin{bmatrix} \widehat{A}_{k\ell}^{[i]} \\ \widehat{A}_{\ell k}^{[i]} \end{bmatrix} = \left\{ \sum_{m=1}^M \begin{bmatrix} \widehat{\mathbf{r}}_{\ell\ell m}^T \mathbf{q}_{k\ell m}^{[i]} & \widehat{\mathbf{r}}_{k\ell m}^T \mathbf{q}_{k\ell m}^{[i]} \\ \widehat{\mathbf{r}}_{k\ell m}^T \mathbf{q}_{k\ell m}^{[i]} & \widehat{\mathbf{r}}_{k\ell m}^T \mathbf{q}_{k\ell m}^{[i]} \end{bmatrix} \right\}^{-1} \sum_{m=1}^M \begin{bmatrix} [\mathbf{q}_{k\ell m}^{[i]}]^T \widehat{\mathbf{r}}_{k\ell m} \\ [\mathbf{q}_{\ell k m}^{[i]}]^T \widehat{\mathbf{r}}_{k\ell m} \end{bmatrix}. \quad (2)$$

In (2), $\widehat{\mathbf{r}}_{k\ell m}$ is a short-hand notation for $\widehat{\mathbf{r}}_{k\ell m}(\mathbf{V}^{[i]})$ and $\mathbf{q}_{k\ell m}^{[i]}$ is derived out of $\widehat{\mathbf{r}}_{k\ell m}(\mathbf{V}^{[i]})$. The algorithm remains asymptotically efficient for the same reason as WASOBI and Block WASOBI; it is only faster.

Computer simulations show that the proposed iterations in (2) has the same, nearly quadratic convergence that is inherent to the Gauss iteration in [13], see the simulation section. We found useful to initialize the algorithm by the outcome of the unweighted AJD algorithm named UWEDGE[13].

3. THE CRB EXPRESSION

A common measure for the performance of BSS algorithms based on an estimate $\widehat{\mathbf{V}}$ of the unmixing matrix is the ISR matrix, whose elements are defined as

$$\text{ISR}_{k\ell} \triangleq E \left[\frac{|(\widehat{\mathbf{V}}\mathbf{A})_{k\ell}|^2}{|(\widehat{\mathbf{V}}\mathbf{A})_{kk}|^2} \right] \frac{\bar{R}_\ell[0]}{\bar{R}_k[0]} \quad k, \ell = 1, \dots, K, k \neq \ell, \quad (3)$$

where $\bar{R}_k[\tau]$ denotes the k -th source's average correlation function (average over all frames of stationarity). $\text{ISR}_{k\ell}$ is essentially the relative mean square residual presence of the ℓ -th source in the estimated k -th source.

The CRB expression for the ISR matrix elements to be derived in this section is an extension of the CRB derived in [5] for one block. It was shown in [5, 12] that the Fisher

information matrix (FIM) for estimating \mathbf{V} for given mixing matrix \mathbf{A} is

$$\mathbf{F}_{\mathbf{V}}(\mathbf{A}) = (\mathbf{A} \otimes \mathbf{I})\mathbf{F}_{\mathbf{V}}(\mathbf{I})(\mathbf{A}^T \otimes \mathbf{I}), \quad (4)$$

where $\mathbf{F}_{\mathbf{V}}(\mathbf{I})$ denotes the FIM when the mixing matrix is $\mathbf{A} = \mathbf{I}$ (with the same AR parameters). The matrix $\mathbf{F}_{\mathbf{V}}(\mathbf{I})$ has elements

$$-E \left[\frac{\partial^2 L_{\mathbf{A}}(\mathbf{V})}{\partial \mathbf{V}_{st} \partial \mathbf{V}_{kl}} \right] \Big|_{\mathbf{A}=\mathbf{I}} = N_B \cdot \begin{cases} 1 + \phi_{kk} & k=t=\ell=s \\ 1 & k=t \neq \ell=s \\ \frac{\sigma_k^2}{\sigma_k^2} \phi_{\ell k} & \ell=t \neq s=k \\ 0 & \text{otherwise} \end{cases} \quad (5)$$

where $L_{\mathbf{A}}(\mathbf{V})$ is the log-likelihood function for the problem and

$$\phi_{k\ell} \triangleq \frac{1}{\sigma_k^2} \sum_{i,j=0}^{p_\ell} a_{\ell,i} a_{\ell,j} R_k[i-j]. \quad (6)$$

In the block-stationary model, each $\phi_{k\ell}$ will have still another index m which specifies the signal interval. The final Fisher information matrix for estimating \mathbf{V} will be given as a sum of the information matrices corresponding to individual frames. Like in [5], the resulting FIM is a diagonal matrix except for the cross terms on the $(\mathbf{V}_{k\ell}, \mathbf{V}_{\ell k})$ - related entries,

$$N_B M \begin{bmatrix} \bar{\phi}_{k\ell} & 1 \\ 1 & \bar{\phi}_{\ell k} \end{bmatrix}$$

where

$$\bar{\phi}_{k\ell} = \frac{1}{M} \sum_{m=1}^M \frac{\sigma_{km}^2}{\sigma_{\ell m}^2} \phi_{k\ell m}$$

and σ_{km}^2 is the variance of the innovation sequence of the k -th source in the m -th interval. The CRB-induced bound on the ISR in (3) is then given as

$$\text{ISR}_{k\ell} \geq \frac{1}{N_B M} \cdot \frac{\bar{\phi}_{k\ell}}{\bar{\phi}_{k\ell} \bar{\phi}_{\ell k} - 1} \cdot \frac{\bar{R}_\ell[0]}{\bar{R}_k[0]}. \quad (7)$$

Since the algorithm is asymptotically efficient, an approximate equality holds in (7). The above expression can be used as an estimate of the separation performance, provided that the theoretical quantities $\bar{\phi}_{k\ell}$ and $\bar{R}_k[0]$ are replaced by their empirical counterparts.

4. SIMULATIONS

First we have tested the algorithm on data that exactly obey the assumed model. We considered three signals, each composed of two stationary blocks. The signal in each block was a first order Gaussian AR process of the length $N_B = 1000$ samples. The first signal had the same spectral shape in each block, having AR coefficients $(1, -0.7\rho)$, where ρ was a free parameter, but the data in the former block had 4 times larger

variance than data in the latter block. The second signal had the same AR coefficients and was stationary. Therefore, these two signals would not be separable by methods that rely on the spectral diversity only. The separation is possible only thanks to unequal variance in the two blocks. The third signal had the same AR coefficients in its former block, but different in the latter block, namely $(1, 0.7\rho)$. The variance in both blocks of the third signal was the same. The same profile of the variance of the second and the third signals makes them non-separable by methods that rely on power-nonstationarity alone. For $\rho \rightarrow 0$, the second and the third signals become non-separable, only the first signal can be separated from the mixture thanks to its non-stationarity.

The data were generated independently in 500 trials, mixed by a random square matrix with independent $\mathcal{N}(0, 1)$ -distributed elements and demixed again. The resultant total (row) ISR for separation of each of the three signals is shown in Figure 1 as a function of the parameter ρ . We can see that the proposed algorithm, BARBI, allows separation of all three signals with $\text{ISR} < -20$ dB, if $\rho > 0.2$. The figures also show performance of the algorithm by Pham [7] with two blocks and four frequencies. We note that the estimate of the first signal is not as good as that of BARBI. It also can be seen that BARBI attains the corresponding CRB, except for cases with small ρ .

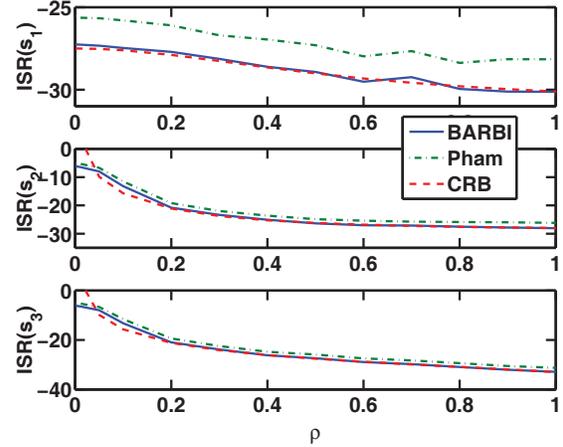


Fig. 1. Average ISR in dB obtained in separation of three artificial blockwise AR(1) signals versus parameter ρ .

The second example consists in separation of a linear instantaneous mixture of natural speech data. We used speech utterances of $d = 15$ different speakers (male and female), sampled at 8000Hz. In each trial, each source was taken as $N = 6000$ samples of speech by the respective speakers, starting at random times. The sources were mixed using a random matrix and blindly separated by the proposed algorithm with different parameters. The result for Pham's algorithm are shown for the same number of blocks and for four frequencies. We can see that the best separation is obtained

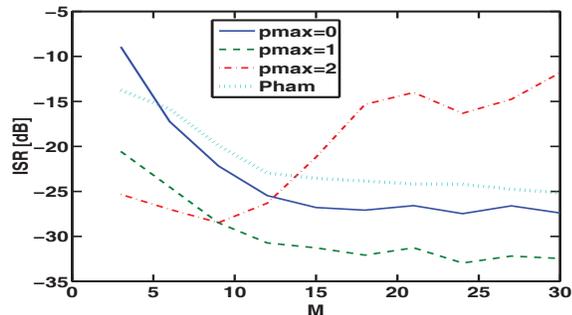


Fig. 2. Average ISR obtained in 100 independent trial in separation of 15 natural speech signals.

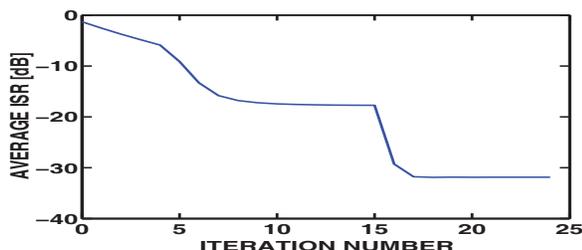


Fig. 3. Learning curve of the UWEDGE/BARBI algorithm.

for the first-order AR model, the method is not very sensitive to the number of blocks. Optimum number of blocks lies somewhere between 18 and 30. Figure 3 shows the typical learning curve of the algorithm for the second example, 25 blocks (frames) and the first AR order. The first 15 iterations are in the initial separation procedure, UWEDGE, the additional iterations are that of BARBI. It reveals fast convergence of both algorithms.

In the last example we separated an overdetermined noisy mixture of speech signals. We used the same set of speech signals, but the mixing matrix was of the size 100×15 , again with $\mathcal{N}(0, 1)$ -distributed elements. A white Gaussian noise of the variance $10^{-\text{SNR}/10}$ was added to the mixture for varying SNR. Figure 4 shows the average empirical ISR of the estimated speech signals versus the SNR, computed for our BARBI with the first-order AR model and 25 blocks, and of the algorithm SOBI-RO [1] that estimates only interesting sources. It is shown that BARBI outperforms the latter algorithm even at low SNR. One run of BARBI takes about 17 s on an ordinary PC with a 3 GHz processor, while SOBI-RO takes about 2 s.

5. CONCLUSIONS

The proposed algorithm, BARBI, appears to be fast and robust, and if the data obey the assumed block-AR model, also asymptotically efficient. For separation of speech signals, the

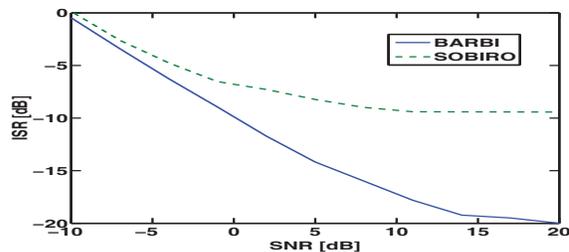


Fig. 4. Average ISR in separating overdetermined noisy mixture of the speech data versus the SNR.

first-order model usually gives the best separation results.

6. REFERENCES

- [1] A. Belouchrani, and A. Cichocki, "Robust whitening procedure in blind source separation context", *Electronics Letters*, Vol. 36, No. 24, 2000, pp. 2050-2053.
- [2] H. Buchner, R. Aichner, and W. Kellermann, "A generalization of blind source separation algorithms for convolutive mixtures based on second-order statistics", *IEEE Tr. on Audio, Speech, and Language Processing*, vol. 13, no. 1, Jan. 2005, pp. 120 - 134.
- [3] J.-F. Cardoso, "The three easy routes to independent component analysis; contrasts and geometry," *Proc. ICA 2001, San Diego, Dec. 2001*.
- [4] S. Choi, A. Cichocki, and A. Belouchrani, "Second order nonstationary source separation," *Journal of VLSI Signal Processing*, vol. 32, no. 1-2, pp. 93-104, Aug. 2002.
- [5] E. Doron, A. Yeredor and P. Tichavský, "Cramér-Rao Lower Bound for Blind Separation of Stationary Parametric Gaussian Sources", *IEEE Signal Processing Letters*, vol. 14, no.6, pp. 417-420, June 2007.
- [6] U. Manmohtri and P.A. Naylor, "A Class of Frobenius Norm-Based Algorithms Using Penalty Term and Natural Gradient for Blind Signal Separation", *IEEE Tr. on Audio, Speech, and Language Processing*, vol. 16, no. 6, Aug. 2008, pp.1181 - 1193.
- [7] D.-T. Pham, "Exploiting Source Non Stationary and Coloration in Blind Source Separation", *Proceeding of DSP 2002 Conference*, Santorini, Greek, July 2002.
- [8] D.-T. Pham and Ph. Garat, "Blind Separation of Mixtures of Independent Sources Through a Quasi Maximum Likelihood Approach," *IEEE Trans. Signal Processing*, vol. 45, no.7, pp. 1712-1725, July 1997.
- [9] P. Tichavský, E. Doron, A. Yeredor and J. Nielsen, "A Computationally Affordable Implementation of An Asymptotically Optimal BSS Algorithm for AR Sources", *Proc. EUSIPCO-2006*, Florence, Italy, September 4-8, 2006.
- [10] P. Tichavský, matlab code for UWEDGE, WEDGE, WASOBI, and BARBI, <http://si.utia.cas.cz/Tichavsky.html>.
- [11] P. Tichavský, A. Yeredor and J. Nielsen, "A Fast Approximate Joint Diagonalization Algorithm Using a Criterion with a Block Diagonal Weight Matrix", *ICASSP 2008*, Las Vegas, March 30-April 4, 2008, pp. 3321-3324.
- [12] P. Tichavský, Z. Koldovsky, and E. Oja, "Corrections to "Performance Analysis of the FastICA Algorithm and Cramér-Rao Bounds for Linear Independent Component Analysis", TSP 04/06", *IEEE Tr. Signal Processing*, vol. 56, no. 4, pp. 1715-1716, April 2008.
- [13] P. Tichavský and A. Yeredor, "Fast Approximate Joint Diagonalization Incorporating Weight Matrices", *IEEE Tr. on Signal Processing*, to appear.
- [14] A. Yeredor, "Blind separation of Gaussian sources via second-order statistics with asymptotically optimal weighting," *IEEE Signal Processing Letters*, vol. 7, pp. 197-200, July 2000.