AN ADAPTIVE PARAUNITARY APPROACH FOR BLIND EQUALIZATION OF ALL EQUALIZABLE MIMO CHANNELS

Alper Tunga Erdogan

Koc University, Istanbul, Turkey,
Email: alperdogan@ku.edu.tr

ABSTRACT

We introduce a novel adaptive paraunitary approach to be used for the blind deconvolution of all deconvolvable MIMO mixing systems with memory. The proposed adaptive approach is based on the use of alternating projections technique for the enforcement of the paraunitary constraint. The use of this approach enables extension of various instantaneous Blind Source Separation (BSS) approaches to handle the convolutive BSS case. Three such methods, namely FastICA, Multi User Kurtosis and BSS for Bounded Magnitude signals are provided to illustrate the use of this approach.

1. INTRODUCTION

In the Blind Source Separation area, a current focus of research is the extension of the methods developed for the instantaneous BSS problems to the convolutive BSS separation problems. However, this extension is not trivial due to the paraunitary constraint that arises in implicit or explicit Higher Order Statistics (HOS) optimization problems targetting to resolve the phase ambiguity (due to mixing in space and time), which can not be resolved using only second order statistics (SOS). The goal of this article is to provide a relatively simple and intuitive approach for the incorporation of the paraunitary constraint into the BSS approaches.

One key result in extending BSS procedures to convolutive mixuters is the theorem due to Inouye et. al. in [1], which states that a MIMO channel $\mathbf{H}(z)$ is blindly equalizable (or deconvolvable) if and only if it can be factorized as

$$\mathbf{H}(z) = \mathbf{H}_I(z)\mathbf{H}_P(z), \tag{1}$$

where $\mathbf{H}_I(z)$ is irreducible and $\mathbf{H}_P(z)$ is paraunitary. This theorem naturally inspires the following two step procedure for the MIMO Blind Deconvolution: SOS based whitening procedure to compensate the irreducible component of the mixing transfer matrix followed by an HOS based method to identify and/or compensate the paraunitary component.

There is a considerable amount of research done related to the whitening step which are based on the linear prediction (see for example [2, 3]). The area that probably needs more attention is the second step which is the design of algorithms for the blind deconvolution of paraunitary systems.

If we look at the pioneering work done in this area: Matsuoka et.al. introduced Cayley Transform based approach to convert paraunitary constraint into para-skew-hermitian constraints [4]. In [5] De Lathauwer et. al. used the parametrization based on the decomposition of paraunitary transfer matrices in terms of simple degree one paraunitary blocks and unitary matrices. Recently, Douglas et.al. introduced an iterative correction procedure to be applied after the FastICA gradient update that targets to preserve the paraunitariness of the search transfer matrix [3].

In this article we introduce another iterative procedure to enforce the paraunitariness constraint and apply it to different instantaneous BSS approaches to extend their use to convolutive BSS case. The approach is based on the fact that the convolution matrices of the paraunitary matrices have orthonormal rows . Therefore the iterative procedure is based on alternating projections on the set of matrices with orthonormal rows and the set of block convolution matrices.

The organization of the article is as follows. In Section 2 we introduce the convolutive BSS setup that we use throughout the article. In Section 3 the proposed adaptive paraunitary approach and its application to different BSS criteria is provided. A simulation example is provided in Section 4. Finally, Section 5 is the conclusion.

2. CONVOLUTIVE BSS SETUP

The components of the blind source separation setup that we consider throughout the article:

- $s_1(k), s_2(k), \ldots, s_p(k)$ are source signals where it is assumed that they are all i.i.d with zero mean and unity variance (without loss of generality), and mutually independent of each other.

- The source signals are mixed by a MIMO system with a $q \times p$ transfer matrix $\mathbf{H}(z)$, which has q outputs denoted by $y_1(k), y_2(k), \ldots, y_q(k)$. In z-transform domain, we

This work was supported in part by TUBITAK Career Award, Contract No:104E073

can write

$$\underbrace{\left[\begin{array}{cccc} y_1(z) & y_2(z) & \dots & y_q(z) \end{array}\right]^T}_{\mathbf{y}(z)} = \mathbf{H}(z) \underbrace{\left[\begin{array}{cccc} s_1(z) & s_2(z) & \dots & s_p(z) \end{array}\right]^T}_{\mathbf{s}(z)}.$$

We assume that $\mathbf{H}(z)$ is an equalizable channel such that it can be written as in (1). The problem boils down to an instantaneous BSS problem after the whitening if the paraunitary matrix $\mathbf{H}_P(z)$ is equal to a unitary matrix or to a transfer function of the form

$$\mathbf{H}_P(z) = diag(z^{d_1}, \dots, z^{d_p})\mathbf{\Phi} \tag{2}$$

where the diagonal matrix corresponds to delaying of sources with integer delays d_k and $\mathbf{\Phi}$ is a unitary mapping.

- Separator $\mathbf{W}(z)$ is a $q \times p$ transfer matrix, where the separator output can be written as

$$\mathbf{o}(z) = \mathbf{W}^T(z)\mathbf{y}(z). \tag{3}$$

We assume that the $\mathbf{W}(z)$ is decomposed as $\mathbf{W}(z) = \mathbf{W}_{pre}(z)\mathbf{\Theta}(z)$. Here $\mathbf{W}_{pre}(z)$ is the $q \times p$ whitening matrix and $\mathbf{x}(z) = \mathbf{W}_{pre}^T(z)\mathbf{y}(z)$ is the whitened mixtures. Since $\mathbf{W}_{pre}^T(z)\mathbf{H}_I(z) = \mathbf{\Phi}$ where $\mathbf{\Phi}$ is a unitary matrix, the effective mapping between $\mathbf{x}(z)$ and $\mathbf{s}(z)$ is equivalent to

$$\begin{aligned} \mathbf{x}(z) &= \mathbf{W}_{pre}^T(z)\mathbf{y}(z) & (4)\\ &= \mathbf{W}_{pre}^T(z)\mathbf{H}_I(z)\mathbf{H}_P(z)\mathbf{s}(z) & (5)\\ &= \mathbf{\Phi}\mathbf{H}_P(z)\mathbf{s} = \mathbf{H}'_P(z)\mathbf{s}(z) & (6) \end{aligned}$$

The goal of adaptive paraunitary stage is to train the paraunitary transfer function $\mathbf{\Theta}(z) = \sum_{k=0}^{\mathcal{L}-1} \Theta_k z^k$ (whose order should be greater than or equal to $\mathbf{H}'_P(z)$) where the goal is to compensate the effective channel $\mathbf{H}'_P(z)$, i.e., to obtain a $\mathbf{\Theta}(z)$ for which

$$\mathbf{\Theta}^T(z)\mathbf{H}'_P(z) = diag(e^{j\phi_1}z^{d_1}, \dots, e^{j\phi_p}z^{d_p})\mathbf{E} \tag{7}$$

where d_k's and ϕ_k's are delays and phase ambiguities of the recovered sources respectively, and $\mathbf{E}$ is a permutation matrix representing permutation ambiguity.

We introduce following matrices corresponding to the transfer function$\mathbf{\Theta}(z)$ and its adaptation:

- The cascaded impulse response matrix:

$$\Upsilon = \left[\begin{array}{cccc} \mathbf{\Theta}_0^T & \mathbf{\Theta}_1^T & \dots & \mathbf{\Theta}_{\mathcal{L}-1}^T \end{array}\right]^T, \tag{8}$$

- The convolution matrix:

$$\mathcal{T}_{\mathcal{L}}(\mathbf{\Theta}(z)) = \left[\begin{array}{cccccccc} \mathbf{\Theta}_0 & \mathbf{\Theta}_1 & \mathbf{\Theta}_2 & \dots & \mathbf{\Theta}_{\mathcal{L}-1} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \mathbf{\Theta}_0 & \mathbf{\Theta}_1 & \dots & \mathbf{\Theta}_{\mathcal{L}-2} & \mathbf{\Theta}_{\mathcal{L}-1} & \dots & \mathbf{0} \\ \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \\ \mathbf{0} & \dots & \mathbf{0} & \mathbf{\Theta}_0 & \dots & \dots & \dots & \mathbf{\Theta}_{\mathcal{L}-1} \end{array}\right]$$

which has size $\mathcal{L}p \times (2\mathcal{L}-1)p$.

- Input data matrix:

$$\mathbf{X} = \left[\begin{array}{cccc} \mathbf{x}(0) & \mathbf{x}(1) & \dots & \mathbf{x}(\Omega-1) \\ \mathbf{x}(-1) & \mathbf{x}(0) & \dots & \mathbf{x}(\Omega-1) \\ \vdots & \vdots & \dots & \vdots \\ \mathbf{x}(-\mathcal{L}+1) & \mathbf{x}(-\mathcal{L}+2) & \dots & \mathbf{x}(\Omega-\mathcal{L}) \end{array}\right],$$

where Ω is the output window size.

- Output matrix

$$\mathbf{O} = \left[\begin{array}{ccc} \mathbf{o}(0) & \dots & \mathbf{o}(\Omega-1) \end{array}\right] = \Upsilon^T\mathbf{X}.$$

3. ADAPTIVE PARAUNITARY APPROACH

We can formulate a general adaptive paraunitary approach as the optimization problem in the form

$$\begin{array}{ll} \text{optimize} & \mathcal{J}(\mathbf{\Theta}(z), \mathbf{X}) \\ \text{s. t.} & \mathbf{\Theta}(z)\mathbf{\Theta}^H(z^{-*}) = \mathbf{I}. \end{array}$$

where $\mathcal{J}$ is the cost function to be minimized or maximized. A typical gradient (or subgradient) based adaptation procedure consists of following two steps:

Step 1 $\underline{\mathbf{\Upsilon}}^{(i+1)} = \mathbf{\Upsilon}^{(i)} + \mathbf{C}^{(i)}$

Step 2 $\mathbf{\Upsilon}^{(\mathbf{i+1})} = \mathcal{M}_{\mathbf{P}}(\underline{\mathbf{\Upsilon}}^{(i+1)})$

where Step 1 is the gradient (or subgradient) based correction on $\mathbf{\Upsilon}$ using correction matrix $\mathbf{C}^{(i)}$ and Step 2 is the enforcement of the paraunitary constraint through the mapping $\mathcal{M}_P$.

Step 1 can be easily generalized from instantaneous BSS approaches. The critical step is Step 2 where we need an efficient method for the enforcement of paraunitary constraint.

Our approach for the design of mapping $\mathcal{M}_P$ is inspired by the basic property of paraunitary transfer matrices that is described by the following theorem [6]:

Theorem 1: A $p \times p$ transfer function $\mathbf{\Theta}(z)$ of order $\mathcal{L}-1$ is paraunitary if and only if the $\mathcal{L}p \times (2\mathcal{L}-1)p$ convolution matrix $\mathcal{T}_{\mathcal{L}}(\mathbf{\Theta}(z))$ has orthonormal rows.

Therefore, the mapping $\mathcal{M}_P$ should target at obtaining an $\mathbf{\Upsilon}^{(i+1)}$ with a corresponding convolution matrix with orthonormal rows. Taking this fact as the central point, we define $\mathcal{M}_P$ based on the alternative projections between the following two sets which are subsets of $\mathcal{C}^{\mathcal{L}p \times (2\mathcal{L}-1)p}$:

$$\mathcal{A}_1 = \{\mathbf{A} \in \mathcal{C}^{\mathcal{L}p \times (2\mathcal{L}-1)p} \mid \mathbf{A}\mathbf{A}^H = \mathbf{I}\},$$

and

$$\mathcal{A}_2 = \{\mathbf{A} \in \mathcal{C}^{\mathcal{L}p \times (2\mathcal{L}-1)p} \mid \mathbf{A} \text{ is a } \mathcal{L}p \times (2\mathcal{L}-1)p \text{ block convolution matrix with } p \times p \text{ blocks}\}$$

Note that the intesection of these two sets

$$\mathcal{A} = \mathcal{A}_1 \bigcap \mathcal{A}_2 \tag{9}$$

would be the set of block convolution matrices with orthonormal rows which is the set of block convolution matrices corresponding to the set of paraunitary operators. Therefore, the task of projecting $\underline{\mathbf{\Theta}}^{(i+1)}(z)$ (transfer function corresponding to $\underline{\mathbf{\Upsilon}}^{(i+1)}$) to set of paraunitary matrices is equivalent to projecting $\mathcal{T}_{\mathcal{L}}(\underline{\mathbf{\Theta}}^{(i+1)}(z))$ to the set $\mathcal{A}$.

Using the fact that $\mathcal{A}$ is the intersection of $\mathcal{A}_1$ and $\mathcal{A}_2$, we can apply alternating projections between the the sets $\mathcal{A}_1$ and $\mathcal{A}_2$ for the goal of enforcing paraunitary constraints. We should note that although the set $\mathcal{A}_2$ is convex, $\mathcal{A}_1$ is not, and therefore, this wouldn't be a POCS algorithm The reason for this choice is that as there is no known explicit method for the projection to $\mathcal{A}$, however, we have explicit methods for the projections to the sets $\mathcal{A}_1$ and $\mathcal{A}_2$ as outlined by the following theorems:

Theorem 2 (*Projection to set* $\mathcal{A}_1$): Let $\mathbf{B} \in \mathcal{C}^{\mathcal{L}p \times (2\mathcal{L}-1)p}$ with a singular value decomposition

$$\mathbf{B} = \mathbf{U} \left[\begin{array}{cc} \mathbf{\Sigma} & \mathbf{0} \end{array} \right] \mathbf{V}^H \tag{10}$$

where $\mathbf{U} \in \mathcal{C}^{\mathcal{L}p \times \mathcal{L}p}$, $\mathbf{V} \in \mathcal{C}^{(2\mathcal{L}-1)p \times (2\mathcal{L}-1)p}$ are unitary matrices, and $\mathbf{\Sigma} \in \Re^{\mathcal{L}p \times \mathcal{L}p}$ is a diagonal matrix with non-diagonal entries. Then $\mathbf{A} \in \mathcal{A}_1$ that minimizes $\|\mathbf{B} - \mathbf{A}\|_F^2$ is given by

$$\mathbf{A} = \mathbf{U} \left[\begin{array}{cc} \mathbf{I} & \mathbf{O} \end{array} \right] \mathbf{V}^H. \tag{11}$$

Proof: See e.g. [7].

Theorem 3 (*Projection to set* $\mathcal{A}_2$): Let $\mathbf{B} \in \mathcal{C}^{\mathcal{L}p \times (2\mathcal{L}-1)p}$. Then $\mathbf{A} \in \mathcal{A}_2$ that minimizes $\|\mathbf{B} - \mathbf{A}\|_F^2$ is given by

$$\mathbf{A} = \mathcal{T}_{\mathcal{L}}(\mathbf{C}(z)), \tag{12}$$

where $\mathbf{C}(z) = \sum_{k=0}^{\mathcal{L}-1} \mathbf{C}_k z^k$ and

$$\begin{aligned} \mathbf{C}_k &= \frac{1}{\mathcal{L}} \sum_{l=0}^{\mathcal{L}-1} \mathbf{B}_{1+lp:p+lp,(k-1)*p+1+lp:(k-1)*p+p+lp} \\ & k = 0, \ldots, \mathcal{L}-1. \end{aligned} \tag{13}$$

Proof: If we consider $\mathbf{A}$ as the convolution matrix of a transfer matrix $\mathbf{C}(z)$ with order $\mathcal{L}-1$, then

$$\begin{aligned} \|\mathbf{B} - \mathbf{A}\|_F^2 &= \|\mathbf{B} - \mathcal{T}_{\mathcal{L}}(\mathbf{C}(z))\|_F^2 \\ = \sum_{k=0}^{\mathcal{L}-1} \sum_{l=0}^{\mathcal{L}-1} & \|\mathbf{C}_k - \mathbf{B}_{1+lp:p+lp,(k-1)*p+1+lp:(k-1)*p+p+lp}\|_F^2 \\ & + \text{terms independent of } \mathbf{C}_k\text{'s}, \end{aligned}$$

which is clearly minimized by $\mathbf{C}_k$'s in (13).

As a result, we can outline the method based on alternating projections as follows:

- let $\mathcal{P}_{\mathcal{A}_1}$, $\mathcal{P}_{\mathcal{A}_2}$ denote the projection operators defined by Theorem 3 and Theorem 4 respectively.
- let $\bar{k}$ be the maximum number of iterations.

Algorithm

S0. Set $k = 1, \mathbf{B}^{(0)} = \mathcal{T}_{\mathcal{L}}(\underline{\mathbf{\Theta}}^{(i+1)}(z))$

S1. $\underline{\mathbf{B}}^{(k)} = \mathcal{P}_{\mathcal{A}_1}(\mathbf{B}^{(k-1)})$,

S2. $\mathbf{B}^{(k)} = \mathcal{P}_{\mathcal{A}_2}(\underline{\mathbf{B}}^{(k)})$,

S3. Set $k = k + 1$, if $k \leq \bar{k}$ go to S1,

S4. Terminate.

We can provide following specific algorithms based on this two step procedure with alternative projections mapping: (Examples can be extended for other BSS algorithms with similar structure)

BSS for Magnitude Bounded Sources: (MB-BSS) (Introduced in [8] and [9])

$$\begin{aligned} \underline{\mathbf{\Upsilon}}^{(i+1)} &= \mathbf{\Upsilon}^{(i)} - \mu^{(i)} \text{sign}(\Re e\{\mathbf{O}^{(i)}_{m^{(i)},n^{(i)}}\}) \bar{\mathbf{X}}_{:,n^{(i)}} \mathbf{e}^T_{m^{(i)}} \\ \mathbf{\Upsilon}^{(i+1)} &= \mathcal{M}_P\{\underline{\mathbf{\Upsilon}}^{(i+1)}\}, \end{aligned}$$

where $\mathbf{O}^{(i)}_{m^{(i)},n^{(i)}}$ is the output for which the maximum real magnitude is achieved, $\mu^{(i)}$ is the step size and $\mathbf{e}_k$ is the k^{th} standard basis vector. Note that this algorithm exploits, in addition to independence, the boundedness of the source signals assuming

$$\begin{aligned} \sup \Re e\{s_l\} &= \sup Im\{s_l\} = -\inf \Re e\{s_l\} \\ &= -\inf \mathcal{I}m\{s_l\} = M. \end{aligned} \tag{14}$$

which is a good fit for digital communications applications.

Multi User Kurtosis Algorithm (MUK) ([10])

$$\begin{aligned} \underline{\mathbf{\Upsilon}}^{(i+1)} &= \mathbf{\Upsilon}^{(i)} + \mu^{(i)} \text{sign}(K_s) \mathbf{X}_{i,:} \mathcal{O}(i) \\ \mathbf{\Upsilon}^{(i+1)} &= \mathcal{M}_P\{\underline{\mathbf{\Upsilon}}^{(i+1)}\}, \end{aligned}$$

where K_s is the kurtosis of the sources and

$$\mathcal{O}(i) = \left[\begin{array}{ccc} |o_1(i)|^2 o_1(i) & \ldots & |o_p(i)|^2 o_p(i) \end{array} \right].$$

FastICA([11])

$$\begin{aligned} \underline{\mathbf{\Upsilon}}^{(i+1)}_{l,:} &= \frac{1}{N} \sum_{k=1}^{N} ((\bar{\mathbf{X}}_{k,:} \mathbf{O}^{(i)}_{k,:} g(|\mathbf{O}^{(i)}_{k,:}|^2)) - (g(|\mathbf{O}^{(i)}_{k,:}|^2) + \\ & |\mathbf{O}^{(i)}_{k,:}|^2 g'(|\mathbf{O}^{(i)}_{k,:}|^2)) \mathbf{\Upsilon}^{(i)}_{l,:}) \quad l = 1, \ldots, p \\ \mathbf{\Upsilon}^{(i+1)} &= \mathcal{M}_P\{\underline{\mathbf{\Upsilon}}^{(i+1)}\} \end{aligned}$$

where g is the derivative of the nonlinear learning function G and g' is the derivative of g.

4. EXAMPLE

In order to illustrate the use of the proposed method we consider a scenario with 4 sources. We assume a random paraunitary mapping (with order 4) between sources and mixtures (this can be considered as the equivalent setting after

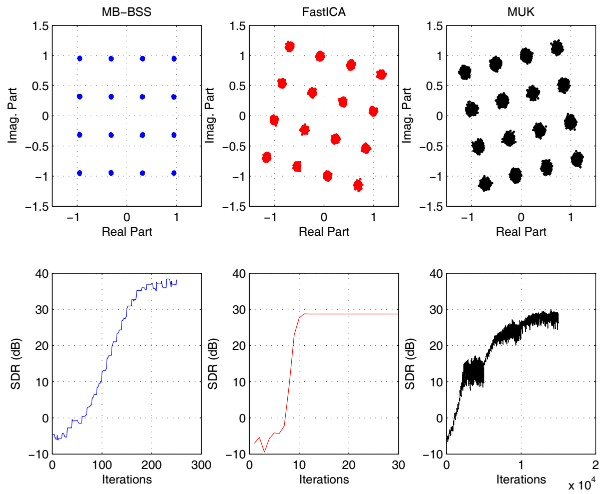

Fig. 1. The SDR convergence curves and the final outputs.

a perfect whitening). Inputs are 16-QAM digital communications signals. Mixtures are corrupted with Gaussian noise and SNR is equal to 40dB. An adaptive paraunitary filter of order 6 is used. Sample Signal to Distortion Ratio (SDR) convergence curves and constellations after the convergence (for one randomly selected output) for the algorithms presented in the previous section are shown in Figure 1. We used a window length of 10000 for Figure 1. SDR curves as a function of available data window size are shown in Figure 2. According to this figure, the proposed approach is successful for all three algorithms. We also note that MB-BSS can achieve higher SNR levels even for short data bursts which can be attributed to the exploitation of the magnitude structure of input signals.

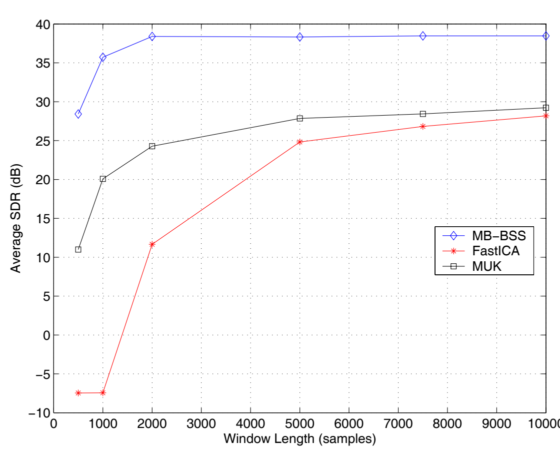

Fig. 2. The achieved SDR levels versus Window Sizes.

5. CONCLUSION

The proposed adaptive paraunitary approach enables the extension of instantaneous BSS approaches to handle convolutive BSS problems. The example provided in the previous section illustrates the successful use of this approach for three different BSS approaches.

6. REFERENCES

[1] Yuijro Inouye and Ruey-Wen Liu, "A system-theoretic foundation for blind equalization of an FIR MIMO channel system," *IEEE Trans. on CAS-I: Fund. Th. and Apps.*, vol. 49, no. 4, pp. 425–436, April 2002.

[2] Y. Inouye, "Modeling of multichannel time series and extrapolation of matrix-valued autocorrelation sequences," *IEEE Trans. Acoust. Speech Signal Processing*, vol. 31, pp. 45–55, February 1983.

[3] S.C. Dogulas, H. Sawada, and S. Makino, "A spatio-temporal FastICA algorithm for separating convolutive mixtures," *IEEE ICASSP 2005, Philadelphia*, vol. 5, pp. 165–168, March 2005.

[4] K. Matsuoka, M. Ohata, and T. Tokunari, "A kurtosis based blind separation of sources using cayley transform," *Proceedings of AS-SPCC*.

[5] L. De Lathauwer, B. De Moor, and J. Vandewalle, "An algebraic approach to blind MIMO identification," *Proceedings of ICA2000, Helsinki*, pp. 211–214, June 2000.

[6] P. P. Vaidyanathan and Z. Doganata, "The role of lossless systems in modern digital signal processing: A tutorial," *IEEE Trans. on Educ.*, pp. 181–197, Aug. 1989.

[7] J. A. Tropp, I. S. Dhillon, R. W. Heath, and T. Strohmer, "Designing structured tight frames via an alternating projection method," *IEEE Trans. on Information Theory*, vol. 51, pp. 188–209, January 2005.

[8] Alper T. Erdogan, "A blind equalization approach for all equalizable MIMO convolutive channels," *under preperation*.

[9] Alper T. Erdogan, "A simple geometric blind source separation method for bounded magnitude sources," *IEEE Trans. on Signal Processing, to appear*.

[10] Constantinos Papadias, "Globally convergent blind source separation based on a multiuser kurtosis maximization criterion," *IEEE Trans. on Signal Processing*, vol. 48, pp. 3508–3519, December 2000.

[11] Aapo Hyvärinen and Erkki Oja, "A fast fixed-point algorithm for independent component analysis," *Neural Computation*, vol. 9, no. 7, pp. 1483–1492, October 1997.