

A Convolutive Bounded Component Analysis Framework for Potentially Non-Stationary Independent and/or Dependent Sources

Huseyin A. Inan, *Student Member, IEEE*, Alper T. Erdogan, *Senior Member, IEEE*
Koc University
Istanbul, Turkey

Abstract—Bounded Component Analysis (BCA) is a recent framework which enables development of methods for the separation of dependent as well as independent sources from their mixtures. This article extends a recent geometric BCA approach introduced for the instantaneous mixing problem to the convolutive mixing problem. The article proposes novel deterministic convolutive BCA frameworks for the blind source extraction and blind source separation of convolutive mixtures of sources which allows the sources to be potentially non-stationary. The global maximizers of the proposed deterministic BCA optimization settings are proved to be perfect separators. The article also illustrates that the iterative algorithms corresponding to these frameworks are capable of extracting/separating convolutive mixtures of not only independent sources but also dependent (even correlated) sources in both component (space) and sample (time) dimensions through simulations based on a Copula distributed source system. In addition, even when the sources are independent, it is shown that the proposed BCA approach have the potential to provide improvement in separation performance especially for short data records based on the setups involving convolutive mixtures of digital communication sources.

Index Terms—Bounded Component Analysis, Independent Component Analysis, Convolutive Blind Source Separation, Dependent Source Separation, Finite Support, Frequency-Selective MIMO Equalization.

I. INTRODUCTION

Blind Source Separation (BSS) is a major area of research with a large scope of applications in signal processing and machine learning [1]. BSS aims to extract individual components (or sources) from their mixture samples where there is no, or very limited, prior information about their nature or the mixing process. The blindness which refers to using only the observations with the absence of information on the mixing system is a striking feature which leads to its widespread use. However, the blindness feature also rises difficulties to the BSS problem where the challenge created by the lack of training data and relational statistical information is in general dealt with exploiting some side information/assumptions about the system.

The BSS problems are mostly addressed by means of Independent Component Analysis (ICA) [1]–[3] exploiting the assumption that the original sources are mutually statistically independent. ICA has been the most popular BSS solution approach, due to the applicability of independence assumption in wide span of applications, as well as the mathematical tractability of the corresponding framework. A variety of other BSS methods have also been emerged from different data model assumptions such as time structure (e.g., [4], [5]), sparsity (e.g., [6]) and special constant modulus or finite alphabet structure of communications signals (e.g. [7]–[9]).

In most practical BSS applications the sources take values from a compact set. Puntonet et.al.’s contribution in [10] can be considered as the pioneering work taking advantage of the geometric structure based on such compactness of sources. ICA framework has an important branch where source boundedness has been exploited as an additional asset in addition to the founding independence assumption. An influential work in this branch is by Pham [11] where the mutual information cost function is approximated in terms of quantile function which can be estimated from the order statistics. In this work, Pham show that when the sources are bounded, the proposed cost function can be formulated in terms of the ranges of separator outputs. In a similar context, Cruces and Duran extend the definition of the Renyi’s Entropy in [12], leading to support length minimization to separate sources from their mixtures. Additionally, Vrins et.al. proposed BSS algorithms exploiting range minimization approach in references [13]–[15]. In a similar direction, Erdogan extend the blind equalization approach in [16], [17] and propose BSS algorithms based on infinity norm minimization for bounded source signals in [18], [19], [20]. where the approaches assume peak symmetry regarding the bounded source signals which is later abandoned in [21].

The aforementioned approaches utilize the assumption of source boundedness within the ICA framework. In a recent work [22], Cruces proved that when the sources are known to be bounded, the hypothesis of the statistical independence of sources can be replaced with a weaker domain separability assumption. This new framework, referred as Bounded Component Analysis (BCA), enables development of methods for the separation of independent and/or dependent sources from their mixtures. The weaker domain separability assumption can be stated as follows: (the convex hull of the) the support of the joint density of sources can be written as the

Huseyin A. Inan is with Electrical-Electronics Engineering Department of Koc University, Sariyer, Istanbul, 34450, Turkey. Contact Information: Email: hinan@ku.edu.tr

Alper T. Erdogan is with Electrical-Electronics Engineering Department of Koc University, Sariyer, Istanbul, 34450, Turkey. Contact Information: Email: alper@stanfordalumni.org, Phone:+90 (212) 338 1490, Fax: +90 (212) 338 1548

cartesian product of (the convex hulls of the) supports of the individual source marginals. We note that this is a necessary condition for the independence, however, the independence assumption additionally requires the joint pdf separability on top of domain separability. Therefore, under the boundedness property ICA becomes a special case of BCA and by replacing the independence assumption with the domain separability assumption, it is possible to separate both independent and dependent sources. Hence, BCA provides a more general and flexible framework than ICA when the sources are bounded.

In this new framework, Cruces introduced a blind source extraction algorithm in [22]. A deflationary algorithm is proposed for BCA in [23]. In [24], total output range minimization based BSS approach is positioned as a BCA method for the separation of uncorrelated sources. Furthermore, a stationary point analysis for the proposed algorithms based on symmetrical orthogonalization is provided. More recently, Erdogan introduced a geometric BCA framework and a family of BCA algorithms in [25], which can separate both independent and dependent (even correlated) sources from their mixtures where the mixing system is instantaneous. In this approach, two geometric objects regarding the separator outputs are introduced, i.e., principal hyperellipsoid and bounding hyperrectangle where the separation problem is based on the maximization of the relative sizes of these objects. When the volume is used for the representation of the size of bounding hyperrectangle, a generalized form of Pham's objective in [11], which was derived by modifying the mutual information objective in ICA framework, is obtained. Based on the similar geometric treatment, a convolutive BCA approach for wide sense stationary (dependent or independent) sources was introduced in [26], [27]. A variety of different geometric approaches are introduced for the context of hyperspectral imaging in [28], [29] and [30] by considering a minimum volume of a simplex that circumscribes the data space.

In this article, we extend the instantaneous or memoryless BCA approach introduced in [25] for the convolutive BCA problem. Contrary to the stochastic convolutive framework in [26], [27] based on the stationary source assumption, we propose deterministic frameworks for the convolutive blind source extraction and blind source separation problems which allows the sources to be potentially non-stationary. We point out that the sources could be stationary or non-stationary and we do not exploit non-stationary property of sources. However, the proposed scheme works for both stationary and non-stationary sources. We show that the algorithms corresponding to these frameworks are capable of extracting/separating convolutive mixtures of not only independent sources but also dependent (even correlated) sources where the correlation can be in both space and time dimensions. We can highlight the novel contributions of the article as:

- 1) The article provides convolutive BCA approaches which can be used to generate algorithms that are capable of extracting/separating dependent sources as well as independent sources from their convolutive mixtures.
- 2) The article proposes a set of convolutive BCA objectives which are directly defined in terms of mixture samples rather than some stochastic measures or their sample

based estimates allowing the source samples to be drawn from a potentially non-stationary process. The introduced framework does not exploit any non-stationarity feature, therefore it is applicable to both stationary/non-stationary sources.

- 3) The article proves that the global optima of proposed BCA objectives correspond to perfect extractors/separators.
- 4) The article illustrates the capability of proposed algorithms regarding the extracting/separating convolutive mixtures of dependent (even correlated) sources. The article further compares the performance of proposed algorithms with the state of the art convolutive ICA approaches and through a digital communications example, the article shows the potential for the significant performance improvement offered by the proposed BCA approach, especially for short data records.

The organization of the article is as follows: Section II describes the convolutive BCA setup assumed throughout the article. The blind source extraction approach is provided in Section III and the blind source separation approach is provided in Section IV. We illustrate the separation performances of the BCA algorithms through the numerical examples in Section V. Finally, Section VI is the conclusion.

Notation: Let $\mathbf{A} \in \mathbb{C}^{p \times q}$ and $\mathbf{a} \in \mathbb{C}^{p \times 1}$ be arbitrary. The notation used in the article is summarized in Table I.

Notation	Meaning
$\mathbf{A}_{m,:}$ ($\mathbf{A}_{:,m}$)	m^{th} row (column) of \mathbf{A}
$\Re\{\mathbf{A}\}$ ($\Im\{\mathbf{A}\}$)	The real (imaginary) part of \mathbf{A}
$\ \mathbf{a}\ _r$	Usual r -norm given by $(\sum_{m=1}^p a ^r)^{1/r}$.
$\text{diag}(\mathbf{a})$	Diagonal matrix whose diagonal entries starting in the upper left corner are a_1, \dots, a_p .
$\prod(\mathbf{a})$	$a_1 a_2 \dots a_p$, i.e. the product of the elements of \mathbf{a} .
$\tilde{\mathbf{a}}_M(k)$	Stacks the vectors $\mathbf{a}(k), \mathbf{a}(k-1), \dots, \mathbf{a}(k-M+1)$ into a single Mp size vector $\tilde{\mathbf{a}}_M(k)$.
\mathbf{e}_m	Standard basis vector pointing in the m^{th} direction.

TABLE I
NOTATION USED IN THE ARTICLE.

Indexing: m is used for (source, output) vector components, k is the sample index and i is the algorithm iteration index.

II. CONVOLUTIVE BOUNDED COMPONENT ANALYSIS SETUP

In this article, the components of the convolutive BCA setup that we consider are as follows:

- We consider a deterministic setup with p real sources represented by the vector $\mathbf{s} = [s_1 \ s_2 \ \dots \ s_p]^T$ and we assume that the sources are bounded in magnitude.
- The source signals are mixed by a convolutive MIMO channel and produce the mixtures as:

$$\mathbf{y}(k) = \sum_{l=0}^{L-1} \mathbf{H}(l) \mathbf{s}(k-l), \quad k \in \mathbb{Z},$$

where $\{\mathbf{H}(l); l \in \{0, \dots, L-1\}\}$ are the impulse response coefficients of dimension $q \times p$. We assume that the mixing system is equalizable having order of $L-1$ [31]. We also assume that $q \geq p$, therefore, we consider the (over)determined BSS problem. Defining $\tilde{\mathbf{H}} = [\mathbf{H}(0) \ \mathbf{H}(1) \ \dots \ \mathbf{H}(L-1)]$ as the mixing coefficient matrix and $\tilde{\mathbf{s}}_L(k) = [\mathbf{s}^T(k) \ \mathbf{s}^T(k-1) \ \dots \ \mathbf{s}^T(k-L+1)]^T$, we can also write

$$\mathbf{y}(k) = \tilde{\mathbf{H}}\tilde{\mathbf{s}}_L(k), \quad k \in \mathbb{Z}.$$

- We investigate two related problems corresponding to this scenario:

- **Blind Source Extraction** : The mixtures are passed through an extractor system and produce the single output as:

$$o(k) = \sum_{l=0}^{M-1} \mathbf{w}^T(l)\mathbf{y}(k-l), \quad k \in \mathbb{Z}.$$

where $\{\mathbf{w}(l); l \in \{0, \dots, M-1\}\}$ are the impulse response coefficients of dimension $q \times 1$ and $M-1$ is the order of the extractor system. Defining $\tilde{\mathbf{w}} = [\mathbf{w}^T(0) \ \mathbf{w}^T(1) \ \dots \ \mathbf{w}^T(M-1)]^T$ as the extractor coefficient matrix and $\tilde{\mathbf{y}}_M(k) = [\mathbf{y}^T(k) \ \mathbf{y}^T(k-1) \ \dots \ \mathbf{y}^T(k-M+1)]^T$, we can also write

$$o(k) = \tilde{\mathbf{w}}^T\tilde{\mathbf{y}}_M(k), \quad k \in \mathbb{Z}.$$

- **Blind Source Separation** : In this case, the outputs of the separator system are produced as:

$$\mathbf{o}(k) = \sum_{l=0}^{M-1} \mathbf{W}(l)\mathbf{y}(k-l), \quad k \in \mathbb{Z}.$$

where $\{\mathbf{W}(l); l \in \{0, \dots, M-1\}\}$ are the impulse response coefficients of dimension $p \times q$ and $M-1$ is the order of the separator system. Similarly, we define $\tilde{\mathbf{W}} = [\mathbf{W}(0) \ \mathbf{W}(1) \ \dots \ \mathbf{W}(M-1)]$ as the separator coefficient matrix which yields

$$\mathbf{o}(k) = \tilde{\mathbf{W}}\tilde{\mathbf{y}}_M(k), \quad k \in \mathbb{Z}.$$

- The overall mapping from sources to the outputs:

- **Blind Source Extraction** : The impulse response coefficients of the overall system are represented as $\{\mathbf{g}(l); l \in \{0, \dots, P-1\}\}$ where the dimension is $p \times 1$. Note that,

$$\mathbf{g}^T(k) = \sum_{l=0}^{P-1} \mathbf{w}^T(l)\mathbf{H}(k-l),$$

where $P-1 = L+M-2$ is the order of the overall system. Therefore, the sources $\{\mathbf{s}(k) \in \mathbb{R}^p; k \in \mathbb{Z}\}$ and the single extractor output $\{o(k) \in \mathbb{R}; k \in \mathbb{Z}\}$ are related by

$$o(k) = \sum_{l=0}^{P-1} \mathbf{g}^T(l)\mathbf{s}(k-l), \quad k \in \mathbb{Z}.$$

Defining $\tilde{\mathbf{g}} = [\mathbf{g}^T(0) \ \dots \ \mathbf{g}^T(P-1)]^T$ and $\tilde{\mathbf{s}}_P(k) = [\mathbf{s}^T(k) \ \dots \ \mathbf{s}^T(k-P+1)]^T$, we have $o(k) = \tilde{\mathbf{g}}^T\tilde{\mathbf{s}}_P(k)$ for $k \in \mathbb{Z}$.

- **Blind Source Separation** : In this case, the impulse response coefficients of the overall system are represented as $\{\mathbf{G}(l); l \in \{0, \dots, P-1\}\}$ where the dimension is $p \times p$. Note that,

$$\mathbf{G}(k) = \sum_{l=0}^{P-1} \mathbf{W}(l)\mathbf{H}(k-l),$$

where $P-1 = L+M-2$ is the order of the overall system. Therefore, the sources $\{\mathbf{s}(k) \in \mathbb{R}^p; k \in \mathbb{Z}\}$ and the separator outputs $\{\mathbf{o}(k) \in \mathbb{R}^p; k \in \mathbb{Z}\}$ are related by

$$\mathbf{o}(k) = \sum_{l=0}^{P-1} \mathbf{G}(l)\mathbf{s}(k-l), \quad k \in \mathbb{Z}.$$

Defining $\tilde{\mathbf{G}} = [\mathbf{G}(0) \ \mathbf{G}(1) \ \dots \ \mathbf{G}(P-1)]$, we have $\mathbf{o}(k) = \tilde{\mathbf{G}}\tilde{\mathbf{s}}_P(k)$, for $k \in \mathbb{Z}$.

We assume a finite set of observations corresponding to the mixture samples represented by $Y = \{\mathbf{y}(1), \mathbf{y}(2), \dots, \mathbf{y}(N)\}$. The main goal in BSS problems is to adapt the extractor/separator system based on these observations. Since the mixing channel $\tilde{\mathbf{H}}$ is convolutive having order of $L-1$, the corresponding set of unobservable source samples could be denoted by $S = \{\mathbf{s}(-L+2), \dots, \mathbf{s}(0), \mathbf{s}(1), \mathbf{s}(2), \dots, \mathbf{s}(N)\}$ such that

$$Y = \{\tilde{\mathbf{H}}\tilde{\mathbf{s}}_L(1), \tilde{\mathbf{H}}\tilde{\mathbf{s}}_L(2), \dots, \tilde{\mathbf{H}}\tilde{\mathbf{s}}_L(N)\}.$$

Depending on the system, for a given convolutive extractor channel $\tilde{\mathbf{w}}$ or a separator channel $\tilde{\mathbf{W}}$ having order of $M-1$ with a corresponding convolutive overall channel $\tilde{\mathbf{g}}$ or $\tilde{\mathbf{G}}$ having order of $P-1$, the convolutive nature of channel generates $N-M+1$ outputs and we illustrate the generated set of extractor output $o = \{o(1), o(2), \dots, o(N-M+1)\}$ or separator outputs $O = \{\mathbf{o}(1), \mathbf{o}(2), \dots, \mathbf{o}(N-M+1)\}$ as

$$\begin{aligned} o &= \{\tilde{\mathbf{w}}^T\tilde{\mathbf{y}}_M(M), \tilde{\mathbf{w}}^T\tilde{\mathbf{y}}_M(M+1), \dots, \tilde{\mathbf{w}}^T\tilde{\mathbf{y}}_M(N)\} \\ &= \{\tilde{\mathbf{g}}^T\tilde{\mathbf{s}}_P(M), \tilde{\mathbf{g}}^T\tilde{\mathbf{s}}_P(M+1), \dots, \tilde{\mathbf{g}}^T\tilde{\mathbf{s}}_P(N)\}, \end{aligned}$$

or

$$\begin{aligned} O &= \{\tilde{\mathbf{W}}\tilde{\mathbf{y}}_M(M), \tilde{\mathbf{W}}\tilde{\mathbf{y}}_M(M+1), \dots, \tilde{\mathbf{W}}\tilde{\mathbf{y}}_M(N)\} \\ &= \{\tilde{\mathbf{G}}\tilde{\mathbf{s}}_P(M), \tilde{\mathbf{G}}\tilde{\mathbf{s}}_P(M+1), \dots, \tilde{\mathbf{G}}\tilde{\mathbf{s}}_P(N)\}. \end{aligned}$$

We define the ranges of the sources as

$$\hat{\mathcal{R}}(s_m) = \max(s_m) - \min(s_m),$$

for $m = 1, 2, \dots, p$, where $\max(s_m)(\min(s_m))$ is the maximum (minimum) value of s_m in the set S . It follows that:

$$\mathbf{s}(k) = \Lambda \underline{\mathbf{s}}(k) \quad \text{for } k = -L+2, \dots, N,$$

where $\Lambda = \text{diag}(\hat{\mathcal{R}}(s_1), \hat{\mathcal{R}}(s_2), \dots, \hat{\mathcal{R}}(s_p))$ is the range matrix of \mathbf{s} and $\underline{\mathbf{s}}$ is the corresponding unit range source vector.

In this article, we extend the deterministic instantaneous or memoryless BCA approach introduced in [25] for the

convolutive BCA problem. The objective functions are directly defined in terms of mixture samples rather than some stochastic measures or their sample based estimates which allows the sources to be drawn from either stationary or non-stationary processes. Regarding the bounding hyper-rectangle, defining the set $S_P = \{\tilde{s}_P(M), \tilde{s}_P(M+1), \dots, \tilde{s}_P(N)\}$, we introduce the following assumption:

Assumption: S_P contains the vertices of its (non-degenerate) bounding hyper-rectangle **(A1)**.

III. BLIND SOURCE EXTRACTION

In this section, we first introduce the objective function for the blind source extraction of real signals. We then prove that the global maxima of the introduced objective function correspond to perfect extractors. We provide the iterative algorithm corresponding to the objective function. We conclude with the complex sources extension of the proposed approach.

A. Criterion

We introduce the objective function for the blind source extraction method as

$$J_e(\tilde{\mathbf{w}}) = \frac{\sqrt{\frac{1}{N_1} \sum_{l=1}^{N_1} (o(l) - \hat{\mu}_o)^2}}{\hat{\mathcal{R}}(o)}, \quad (1)$$

where $\hat{\mu}_o = \frac{1}{N_1} \sum_{l=1}^{N_1} o(l)$, $N_1 = N - M + 1$ and $\hat{\mathcal{R}}(o)$ is the range of the single output o . We note that this objective function is deduced from the instantaneous BCA objectives introduced in [25].

We define

$$\hat{\boldsymbol{\mu}}_{\tilde{\mathbf{s}}_P} = \frac{1}{N_1} \sum_{l=M}^N \tilde{\mathbf{s}}_P(l),$$

$$\hat{\mathbf{R}}_{\tilde{\mathbf{s}}_P} = \frac{1}{N_1} \sum_{l=M}^N (\tilde{\mathbf{s}}_P(l) - \hat{\boldsymbol{\mu}}_{\tilde{\mathbf{s}}_P})(\tilde{\mathbf{s}}_P(l) - \hat{\boldsymbol{\mu}}_{\tilde{\mathbf{s}}_P})^T,$$

as the sample covariance matrix of $\tilde{\mathbf{s}}_P$. If sources are stationary, then $\hat{\mathbf{R}}_{\tilde{\mathbf{s}}_P}$ is a block Toeplitz matrix. However, sources are allowed to be non-stationary, therefore, $\hat{\mathbf{R}}_{\tilde{\mathbf{s}}_P}$ may not be a block Toeplitz matrix. Note that this approach does not exploit any structure on $\hat{\mathbf{R}}_{\tilde{\mathbf{s}}_P}$ (i.e., the sources can be stationary or non-stationary). Under the condition $\hat{\mathbf{R}}_{\tilde{\mathbf{s}}_P} \succ 0$, the following theorem shows that maximizing the proposed objective function (1) achieves the goal of blindly extracting a source from a given convolutive mixture for the setup is outlined in Section II.

Theorem 1: Assuming the setup in Section II, $\tilde{\mathbf{H}}$ is equalizable by an FIR extractor matrix of order $M - 1$ and under the validity of **(A1)**, the set of global maxima for J_e in (1) is equal to the set of perfect extractors.

Proof: The proof is provided in Appendix A.

B. Algorithm

In this section, we provide the iterative algorithm corresponding to the optimization setting presented in the previous section.

Rather than maximizing J_e , we maximize its logarithm since with the logarithm operation, we utilize the conversion of ratio expression to the difference expression since it simplifies the update components in the iterative algorithm. Therefore, the new objective function is modified as

$$\begin{aligned} \bar{J}_e(\tilde{\mathbf{w}}) &= \log(J_e(\tilde{\mathbf{w}})) \\ &= \frac{1}{2} \log(\tilde{\mathbf{w}}^T \hat{\mathbf{R}}_{\tilde{\mathbf{y}}_M} \tilde{\mathbf{w}}) - \log(\hat{\mathcal{R}}(o)), \end{aligned} \quad (2)$$

where $\hat{\mathbf{R}}_{\tilde{\mathbf{y}}_M}$ is the sample covariance matrix of $\tilde{\mathbf{y}}_M$.

Note that the derivative of the first part of $\bar{J}_e(\tilde{\mathbf{w}})$ with respect to $\tilde{\mathbf{w}}$ is

$$\frac{\partial \log(\tilde{\mathbf{w}}^T \hat{\mathbf{R}}_{\tilde{\mathbf{y}}_M} \tilde{\mathbf{w}})}{\partial \tilde{\mathbf{w}}} = \frac{2 \hat{\mathbf{R}}_{\tilde{\mathbf{y}}_M} \tilde{\mathbf{w}}}{\tilde{\mathbf{w}}^T \hat{\mathbf{R}}_{\tilde{\mathbf{y}}_M} \tilde{\mathbf{w}}}.$$

Following the similar steps as in [25] for the derivative of $\log(\hat{\mathcal{R}}(o))$, the subgradient based iterative algorithm for maximizing objective function (2) is provided as

$$\begin{aligned} \tilde{\mathbf{w}}^{(i+1)} &= \tilde{\mathbf{w}}^{(i)} + \mu^{(i)} \left(\frac{\hat{\mathbf{R}}_{\tilde{\mathbf{y}}_M} \tilde{\mathbf{w}}}{\tilde{\mathbf{w}}^T \hat{\mathbf{R}}_{\tilde{\mathbf{y}}_M} \tilde{\mathbf{w}}} - \right. \\ &\quad \left. \frac{1}{\hat{\mathcal{R}}(o^{(i)})} (\tilde{\mathbf{y}}_M^{(l^{max(i)})} - \tilde{\mathbf{y}}_M^{(l^{min(i)})}) \right), \end{aligned} \quad (3)$$

where $\mu^{(i)}$ is the step-size at the i^{th} iteration (see e.g., [17] for a discussion on step sizes) and $l^{max(i)}$ ($l^{min(i)}$) is the sample index for which the maximum (minimum) value for the extractor output is achieved at the i^{th} iteration.

C. Extension to Complex Signals

In the complex domain, both mixing and extractor coefficient matrices are complex matrices, i.e., $\tilde{\mathbf{H}} \in \mathbb{C}^{q \times pL}$ and $\tilde{\mathbf{w}} \in \mathbb{C}^{qM \times 1}$. The set of source vectors S is a subset of \mathbb{C}^p , the set of single extractor output o is a subset of \mathbb{C} and the set of mixtures Y is a subset of \mathbb{C}^q .

In this section, we extend the approach introduced in the Section III-A to the complex signals. We modify the objective function as

$$J_{ce}(\tilde{\mathbf{w}}) = \frac{\sqrt{\frac{1}{N_1} \sum_{l=1}^{N_1} (\Re\{o(l)\} - \Re\{\hat{\mu}_o\})^2}}{\hat{\mathcal{R}}(\Re\{o\})}, \quad (4)$$

where $\Re\{\hat{\mu}_o\} = \frac{1}{N_1} \sum_{l=1}^{N_1} \Re\{o(l)\}$ and $\hat{\mathcal{R}}(\Re\{o\})$ is the range of real parts of output o .

We similarly define $\hat{\mathbf{R}}_{\tilde{\mathbf{s}}_P}$ as the sample covariance matrix of $\tilde{\mathbf{s}}_P$ where $\tilde{\mathbf{s}}_P(k) = [\Re\{\mathbf{s}^T(k)\} \quad \Im\{\mathbf{s}^T(k)\} \quad \dots \quad \Re\{\mathbf{s}^T(k-P+1)\} \quad \Im\{\mathbf{s}^T(k-P+1)\}]^T$. Under the condition $\hat{\mathbf{R}}_{\tilde{\mathbf{s}}_P} \succ 0$, the following theorem shows that maximizing the modified objective function (4) achieves the blind source extraction of convolutive mixtures of complex signals.

Theorem 2: Assuming the setup in Section II, $\tilde{\mathbf{H}}$ is equalizable by an FIR extractor matrix of order $M - 1$ and under the validity of **(A1)**, the set of global maxima for J_{ce} in (4) is equal to a subset of perfect extractors.

Proof: The proof is provided in Appendix B.

In the iterative algorithm, we maximize the logarithm of J_{ce} , therefore, the objective function is modified as

$$\begin{aligned} \bar{J}_{ce}(\hat{\mathbf{w}}) &= \log(J_{ce}(\hat{\mathbf{w}})) \\ &= \frac{1}{2} \log\left(\hat{\mathbf{w}}^T \hat{\mathbf{R}}_{\hat{\mathbf{y}}_M} \hat{\mathbf{w}}\right) - \log\left(\hat{\mathcal{R}}(\mathbb{R}\{o\})\right), \end{aligned} \quad (5)$$

where

$$\hat{\mathbf{w}} = [\mathbb{R}\{\mathbf{w}^T(0)\} \quad -\mathbb{I}\{\mathbf{w}^T(0)\} \quad \dots \quad -\mathbb{I}\{\mathbf{w}^T(M-1)\}]^T,$$

and $\hat{\mathbf{R}}_{\hat{\mathbf{y}}_M}$ is the sample covariance matrix of $\hat{\mathbf{y}}_M$. Following similar steps, the iterative algorithm for maximizing objective function (5) is provided as

$$\begin{aligned} \hat{\mathbf{w}}^{(i+1)} &= \hat{\mathbf{w}}^{(i)} + \mu^{(i)} \left(\frac{\hat{\mathbf{R}}_{\hat{\mathbf{y}}_M} \hat{\mathbf{w}}}{\hat{\mathbf{w}}^T \hat{\mathbf{R}}_{\hat{\mathbf{y}}_M} \hat{\mathbf{w}}} - \right. \\ &\quad \left. \frac{1}{\hat{\mathcal{R}}(\mathbb{R}\{o\})^{(i)}} \left(\hat{\mathbf{y}}_M^{(l^{max(i)})} - \hat{\mathbf{y}}_M^{(l^{min(i)})} \right) \right), \end{aligned} \quad (6)$$

where $\mu^{(i)}$ is the step-size at the i^{th} iteration and $l^{max(i)}$ ($l^{min(i)}$) is the sample index for which the maximum (minimum) value of the real part of the extractor output is achieved at the i^{th} iteration. Finally, we can obtain $\tilde{\mathbf{w}}$ from $\hat{\mathbf{w}}$ using a simple transition $\tilde{\mathbf{w}}_{mq+1:(m+1)q} = \hat{\mathbf{w}}_{2mq+1:2(m+1)q-q} - j\hat{\mathbf{w}}_{2(m+1)q-q+1:2mq}$ for $m = 0, 1, \dots, M-1$.

D. Complexity Analysis

The complexity of the introduced algorithm can be specified by 4 parts.

- Calculating the single separator output $o = \tilde{\mathbf{w}}^T \tilde{\mathbf{y}}_M$: The complexity is $\mathcal{O}(qMN)$.
- Calculating the range of real components of the single separator output $\hat{\mathcal{R}}(\mathbb{R}\{o\})$: The complexity is $\mathcal{O}(N)$.
- Calculating the sample covariance matrix $\hat{\mathbf{R}}_{\hat{\mathbf{y}}_M}$: The complexity is $\mathcal{O}(M^2q^2N)$.
- Performing the iterative algorithm (6): The complexity is $\mathcal{O}(M^2q^2)$.

Therefore, the overall complexity can be determined as $\mathcal{O}(M^2q^2N)$. The required number of iterations for the convergence depends on the system settings. We will provide examples in Section V.

IV. BLIND SOURCE SEPARATION

In this section, we first introduce an objective function for the blind source separation of real signals. We then prove that the global maxima of the introduced objective function correspond to the perfect separators. We next provide a family of alternative objective functions. After producing the iterative algorithms corresponding to the introduced objective functions, we conclude with the complex extension of the proposed approaches.

A. Criteria

In order to define the first objective function, we use a similar geometric setting introduced in [25]. Defining the set $O_K = \{\tilde{\mathbf{o}}_K(K), \tilde{\mathbf{o}}_K(K+1), \dots, \tilde{\mathbf{o}}_K(N-M+1)\}$, we introduce the following objects corresponding to the sets of output samples O_K and O :

- $\mathcal{P}(O_K)$: This is the hyper-ellipsoid whose center is given by the sample mean of the set O_K , its principal semiaxes directions are determined by the eigenvectors of the sample covariance matrix $\hat{\mathbf{R}}_{\tilde{\mathbf{o}}_K}$ corresponding to O_K and its principal semiaxes lengths are equal to the principal standard deviations, i.e., the square roots of the eigenvalues of $\hat{\mathbf{R}}_{\tilde{\mathbf{o}}_K}$.
- $\mathcal{B}(O)$: This is the bounding hyper-rectangle which is defined as minimum volume box covering all the samples in O and aligning with the coordinate axes.

The first objective function that we introduce for blind source separation is

$$J_{s1}(\tilde{\mathbf{W}}) = \frac{\left(\sqrt{\det(\hat{\mathbf{R}}_{\tilde{\mathbf{o}}_K})}\right)^{1/K}}{\prod_{m=1}^P \hat{\mathcal{R}}(o_m)}, \quad (7)$$

where

$$\begin{aligned} \hat{\boldsymbol{\mu}}_{\tilde{\mathbf{o}}_K} &= \frac{1}{N_2} \sum_{l=K}^{N_1} \tilde{\mathbf{o}}_K(l), \\ \hat{\mathbf{R}}_{\tilde{\mathbf{o}}_K} &= \frac{1}{N_2} \sum_{l=K}^{N_1} (\tilde{\mathbf{o}}_K(l) - \hat{\boldsymbol{\mu}}_{\tilde{\mathbf{o}}_K})(\tilde{\mathbf{o}}_K(l) - \hat{\boldsymbol{\mu}}_{\tilde{\mathbf{o}}_K})^T, \end{aligned}$$

$N_2 = N_1 - K + 1$ such that $\hat{\mathbf{R}}_{\tilde{\mathbf{o}}_K}$ is the sample covariance matrix of $\tilde{\mathbf{o}}_K$. $\hat{\mathcal{R}}(o_m)$ is the range of the m 'th component of the output vectors in the set O and we choose $K \geq P$ where P is the order of the overall system.

We note that, as defined in [25],

- $\sqrt{\det(\hat{\mathbf{R}}_{\tilde{\mathbf{o}}_K})}$ refers to the scaled volume of principal hyper-ellipse for the extended output vector $\tilde{\mathbf{o}}_K$.
- $\prod_{m=1}^P \hat{\mathcal{R}}(o_m)$ is the volume of the bounding hyper-rectangle for the output vector \mathbf{o} .

Under the condition $\hat{\mathbf{R}}_{\tilde{\mathbf{s}}_{K+P-1}} \succ 0$, the following theorem shows that maximizing the objective function (7) achieves the blind source separation of convolutive mixtures whose setup is outlined in Section II.

Theorem 3: Assuming the setup in Section II, $\tilde{\mathbf{H}}$ is equalizable by an FIR separator matrix of order $M-1$ and under the validity of **(A1)**, the set of global maxima for J_{s1} in (7) is equal to the set of perfect separator matrices.

Proof: The proof is provided in Appendix C.

We can propose different alternatives for the denominator of the objective function (7) (measure of the size of the bounding hyperrectangle for the output vectors). We can choose the length of the main diagonal of the bounding hyperrectangle as a measure of the size instead of its volume. As a result, we obtain a family of alternative objective functions in the form

$$J_{s2,r}(\tilde{\mathbf{W}}) = \frac{\left(\sqrt{\det(\hat{\mathbf{R}}_{\tilde{\mathbf{o}}_K})}\right)^{1/K}}{\|\hat{\mathcal{R}}(\mathbf{o})\|_r^p}, \quad (8)$$

where $r \geq 1$. We provide the results of analysing this family of objective functions, for some special r values (i.e., $r = 1, 2, \infty$) in Appendix D.

B. Algorithms

In this section, we provide the iterative algorithms corresponding to the optimization settings presented in the previous section.

- *Objective Function $J_{s1}(\tilde{\mathbf{W}})$:*

Similar to the approach in blind source extraction, rather than maximizing $J_{s1}(\tilde{\mathbf{W}})$, we maximize its logarithm. Therefore, the new objective function is modified as

$$\begin{aligned} \bar{J}_{s1}(\tilde{\mathbf{W}}) &= \log \left(J_{s1}(\tilde{\mathbf{W}}) \right) \\ &= \frac{1}{2K} \log \left(\det \left(\Gamma_K(\tilde{\mathbf{W}}) \hat{\mathbf{R}}_{\tilde{\mathbf{y}}_{K+M-1}} \Gamma_K(\tilde{\mathbf{W}})^T \right) \right) \\ &\quad - \log \left(\prod_{m=1}^p \hat{\mathcal{R}}(o_m) \right), \end{aligned} \quad (9)$$

where $\hat{\mathbf{R}}_{\tilde{\mathbf{y}}_{K+M-1}}$ is the sample covariance matrix of $\tilde{\mathbf{y}}_{K+M-1}$. Note that the derivative of the first part of $\bar{J}_{s1}(\tilde{\mathbf{W}})$ with respect to $\tilde{\mathbf{W}}$ is

$$\begin{aligned} \frac{\partial \log \left(\det \left(\Gamma_K(\tilde{\mathbf{W}}) \hat{\mathbf{R}}_{\tilde{\mathbf{y}}_{K+M-1}} \Gamma_K(\tilde{\mathbf{W}})^T \right) \right)}{\partial \tilde{\mathbf{W}}} &= \\ 2 \sum_{l=0}^{K-1} \mathbf{A}_{lp+1:(l+1)p, lq+1:(l+M)q} & \end{aligned}$$

where $\mathbf{A} = \left(\Gamma_K(\tilde{\mathbf{W}}) \hat{\mathbf{R}}_{\tilde{\mathbf{y}}_{K+M-1}} \Gamma_K(\tilde{\mathbf{W}})^T \right)^{-1} \Gamma_K(\tilde{\mathbf{W}}) \hat{\mathbf{R}}_{\tilde{\mathbf{y}}_{K+M-1}}$. Following the similar steps as in [25] for the derivative of $\log \left(\prod_{m=1}^p \hat{\mathcal{R}}(o_m) \right)$, the subgradient based iterative algorithm for maximizing objective function (9) is provided as

$$\begin{aligned} \tilde{\mathbf{W}}^{(i+1)} &= \tilde{\mathbf{W}}^{(i)} \\ &+ \mu^{(i)} \left(\frac{1}{K} \sum_{l=0}^{K-1} \mathbf{A}_{lp+1:(l+1)p, lq+1:(l+M)q} - \right. \\ &\quad \left. \sum_{m=1}^p \frac{1}{\mathbf{e}_m^T \hat{\mathcal{R}}(\mathbf{o}^{(i)})} \mathbf{e}_m \left(\tilde{\mathbf{y}}_M(l_m^{max(i)}) - \tilde{\mathbf{y}}_M(l_m^{min(i)}) \right)^T \right), \end{aligned} \quad (10)$$

where $\mu^{(i)}$ is the step-size at the i^{th} iteration and $l_m^{max(i)}$ ($l_m^{min(i)}$) is the sample index for which the maximum (minimum) value for the m^{th} separator output is achieved at the i^{th} iteration.

- *Objective Function $J_{s2,r}(\tilde{\mathbf{W}})$:*

We note that for the family of objective functions (8), the update equation is similar to (10) where the change is in the derivative of logarithm of the denominator depending on the selection of r . For $r = 1, 2$, we can write the

update equation as

$$\begin{aligned} \tilde{\mathbf{W}}^{(i+1)} &= \tilde{\mathbf{W}}^{(i)} + \\ \mu^{(i)} \left(\frac{1}{K} \sum_{l=0}^{K-1} \mathbf{A}_{lp+1:(l+1)p, lq+1:(l+M)q} - \sum_{m=1}^p \right. & \\ \left. \frac{p \hat{\mathcal{R}}_m(\mathbf{o}^{(i)})^{r-1}}{\|\hat{\mathcal{R}}(\mathbf{o}^{(i)})\|_r^r} \mathbf{e}_m \left(\tilde{\mathbf{y}}_M(l_m^{max(i)}) - \tilde{\mathbf{y}}_M(l_m^{min(i)}) \right)^T \right). & \end{aligned}$$

For $r = \infty$, the update equation has the form

$$\begin{aligned} \tilde{\mathbf{W}}^{(i+1)} &= \tilde{\mathbf{W}}^{(i)} + \\ \mu^{(i)} \left(\frac{1}{K} \sum_{l=0}^{K-1} \mathbf{A}_{lp+1:(l+1)p, lq+1:(l+M)q} - \sum_{m \in \mathcal{M}(\mathbf{o}^{(i)})} \right. & \\ \left. \frac{p \beta_m^{(i)}}{\|\hat{\mathcal{R}}(\mathbf{o}^{(i)})\|_\infty} \mathbf{e}_m \left(\tilde{\mathbf{y}}_M(l_m^{max(i)}) - \tilde{\mathbf{y}}_M(l_m^{min(i)}) \right)^T \right) & \end{aligned}$$

where $\mathcal{M}(\mathbf{o}^{(i)})$ is the set of indexes for which the peak range value is achieved, i.e.,

$$\mathcal{M}(\mathbf{o}^{(i)}) = \{m : \hat{\mathcal{R}}_m(\mathbf{o}^{(i)}) = \|\hat{\mathcal{R}}(\mathbf{o}^{(i)})\|_\infty\}, \quad (11)$$

and $\beta_m^{(i)}$ s are the convex combination coefficients.

C. Extension to Complex Signals

In the complex domain, both mixing and separator coefficient matrices are complex matrices, i.e., $\tilde{\mathbf{H}} \in \mathbb{C}^{q \times pL}$ and $\tilde{\mathbf{W}} \in \mathbb{C}^{p \times qM}$. The set of source vectors S and the set of separator outputs O are a subset of \mathbb{C}^p , the set of mixtures Y is a subset of \mathbb{C}^q .

In this section, we extend the approach introduced in the Section IV-A to the complex signals. We modify the first objective function for the blind source separation of complex signals as

$$J_{cs1}(\tilde{\mathbf{W}}) = \frac{\left(\sqrt{\det(\hat{\mathbf{R}}_{\delta_K})} \right)^{1/K}}{\prod_{m=1}^{2p} \hat{\mathcal{R}}(\delta_m)}, \quad (12)$$

where $\hat{\mathbf{R}}_{\delta_K}$ is the sample covariance matrix of δ_K where $\delta_K(k) = [\mathbb{R}\{\mathbf{o}^T(k)\} \quad \mathbb{I}\{\mathbf{o}^T(k)\} \quad \dots \quad \mathbb{R}\{\mathbf{o}^T(k-K+1)\} \quad \mathbb{I}\{\mathbf{o}^T(k-K+1)\}]^T$ and $\prod_{m=1}^{2p} \hat{\mathcal{R}}(\delta_m)$ is the product of ranges of real and imaginary parts of all separator outputs.

Under the condition $\hat{\mathbf{R}}_{\delta_{K+P-1}} \succ 0$, the following theorem shows that maximizing the modified objective function (12) achieves the blind source separation of convolutive mixtures of complex signals.

Theorem 4: Assuming the setup in Section II, $\tilde{\mathbf{H}}$ is equalizable by an FIR separator matrix of order $M-1$ and under the validity of (A1), the set of global maxima for J_{cs1} in (12) is equal to a subset of perfect separator matrices.

Proof: The proof is provided in Appendix E.

In the iterative algorithm, we maximize the logarithm of J_{cs1} , therefore, the first objective function is modified as

$$\begin{aligned} \bar{J}_{cs1}(\tilde{\mathbf{W}}) &= \log \left(J_{cs1}(\tilde{\mathbf{W}}) \right) \\ &= \frac{1}{2K} \log \left(\det \left(\Gamma_{2K}(\tilde{\mathbf{W}}) \hat{\mathbf{R}}_{\tilde{\mathbf{y}}_{K+M-1}} \Gamma_{2K}(\tilde{\mathbf{W}})^T \right) \right) \\ &\quad - \log \left(\prod_{m=1}^{2p} \hat{\mathcal{R}}(\hat{\mathfrak{o}}_m) \right), \end{aligned} \quad (13)$$

where $\tilde{\mathbf{W}} = \begin{bmatrix} \mathbb{R}\{\mathbf{W}_0\} & -\mathbb{I}\{\mathbf{W}_0\} & \dots & \mathbb{R}\{\mathbf{W}_{M-1}\} & -\mathbb{I}\{\mathbf{W}_{M-1}\} \\ \mathbb{I}\{\mathbf{W}_0\} & \mathbb{R}\{\mathbf{W}_0\} & \dots & \mathbb{I}\{\mathbf{W}_{M-1}\} & \mathbb{R}\{\mathbf{W}_{M-1}\} \end{bmatrix}$ and $\hat{\mathbf{R}}_{\tilde{\mathbf{y}}_{K+M-1}}$ is the sample covariance matrix of $\tilde{\mathbf{y}}_{K+M-1}$.

The corresponding iterative update equation of $\mathbf{W}(n)$ for $n = 0, 1, \dots, M-1$ can be written as

$$\begin{aligned} \mathbf{W}^{(i+1)}(n) &= \mathbf{W}^{(i)}(n) \\ &+ \mu^{(i)} \left(\mathbf{C}_{1:p,2nq+1:(2n+1)q} + \mathbf{C}_{p+1:2p,(2n+1)q+1:2(n+1)q} \right. \\ &+ j \left(\mathbf{C}_{p+1:2p,2nq+1:(2n+1)q} - \mathbf{C}_{1:p,(2n+1)q+1:2(n+1)q} \right) - \\ &\left. \sum_{m=1}^{2p} \frac{1}{2\hat{\mathcal{R}}(\hat{\mathfrak{o}}_m^{(i)})} \mathbf{v}_m \left(\tilde{\mathbf{y}}_M^{(l_m^{max(i)})} - \tilde{\mathbf{y}}_M^{(l_m^{min(i)})} \right)^H \right), \end{aligned} \quad (14)$$

where $\mathbf{C} = \frac{1}{2K} \sum_{l=0}^{K-1} \mathbf{F}_{2lp+1:2(l+1)p,2lq+1:2(l+M)q}$, $\mathbf{F} = \left(\Gamma_{2K}(\tilde{\mathbf{W}}) \hat{\mathbf{R}}_{\tilde{\mathbf{y}}_{K+M-1}} \Gamma_{2K}(\tilde{\mathbf{W}})^T \right)^{-1} \Gamma_{2K}(\tilde{\mathbf{W}}) \hat{\mathbf{R}}_{\tilde{\mathbf{y}}_{K+M-1}}$ and $\mathbf{v}_m = \begin{cases} \mathbf{e}_m & m \leq p, \\ i\mathbf{e}_{m-p} & m > p. \end{cases}$ (15)

Similar to the complex extension of J_{s1} , we can extend the J_{s2} family by modifying

$$J_{cs2,r}(\tilde{\mathbf{W}}) = \frac{\left(\sqrt{\det(\hat{\mathbf{R}}_{\hat{\mathfrak{o}}_K})} \right)^{1/K}}{\|\hat{\mathcal{R}}(\hat{\mathfrak{o}})\|_r^{2p}}. \quad (16)$$

The update equation is similar to (14) where the change is in the derivative of logarithm of the denominator depending on the selection of r , e.g.,

- $r = 1, 2$ Case: In this case

$$\begin{aligned} \frac{\partial \log \left(\|\hat{\mathcal{R}}(\hat{\mathfrak{o}})\|_r^{2p} \right)}{\partial \tilde{\mathbf{W}}} &= \\ \sum_{m=1}^{2p} \frac{p \hat{\mathcal{R}}_m(\hat{\mathfrak{o}}^{(i)})^{r-1}}{\|\hat{\mathcal{R}}(\hat{\mathfrak{o}}^{(i)})\|_r^r} \mathbf{v}_m \left(\tilde{\mathbf{y}}_M^{(l_m^{max(i)})} - \tilde{\mathbf{y}}_M^{(l_m^{min(i)})} \right)^H \end{aligned}$$

where \mathbf{v}_m is as defined in (15).

- $r = \infty$ Case: In this case

$$\begin{aligned} \frac{\partial \log \left(\|\hat{\mathcal{R}}(\hat{\mathfrak{o}})\|_r^{2p} \right)}{\partial \tilde{\mathbf{W}}} &= \sum_{m \in \mathcal{M}(\hat{\mathfrak{o}}^{(i)})} \\ \frac{p \beta_m^{(i)}}{\|\hat{\mathcal{R}}(\hat{\mathfrak{o}}^{(i)})\|_\infty} \mathbf{v}_m \left(\tilde{\mathbf{y}}_M^{(l_m^{max(i)})} - \tilde{\mathbf{y}}_M^{(l_m^{min(i)})} \right)^H \end{aligned}$$

where \mathbf{v}_m is as defined in (15),

$$\mathcal{M}(\hat{\mathfrak{o}}^{(i)}) = \{m : \hat{\mathcal{R}}_m(\hat{\mathfrak{o}}^{(i)}) = \|\hat{\mathcal{R}}(\hat{\mathfrak{o}}^{(i)})\|_\infty\},$$

and $\beta_m^{(i)}$ s are the convex combination coefficients.

D. Complexity Analysis

Similarly, we determine the complexity of the introduced algorithms by considering 4 parts.

- Calculating the separator output $\mathbf{o} = \tilde{\mathbf{W}} \tilde{\mathbf{y}}_M$: The complexity is $\mathcal{O}(pqMN)$.
- Calculating the ranges of separator outputs $\hat{\mathcal{R}}(\mathbf{o})$: The complexity is $\mathcal{O}(pN)$.
- Calculating the sample covariance matrix $\hat{\mathbf{R}}_{\tilde{\mathbf{y}}_{K+M-1}}$: The complexity is $\mathcal{O}((K+M-1)^2q^2N)$.
- Performing the iterative algorithm (14): The complexity is $\mathcal{O}(pK(K+M-1)^2q^2)$.

Therefore, the overall complexity can be determined as $\mathcal{O}((K+M-1)^2q^2N) + \mathcal{O}(pK(K+M-1)^2q^2)$. The required number of iterations for the convergence depends on the system settings. We will provide examples in Section V.

V. EXAMPLES

In this section, we illustrate the extraction/separation capability of the proposed algorithms for the convolutive mixtures of both independent and dependent sources.

A. Blind Source Extraction

We first consider the following scenario to illustrate the performance of the proposed blind source extraction algorithm regarding the convolutive mixtures of space-time correlated sources: In order to generate space-time correlated sources, we first generate a samples of a τp size vector, \mathbf{d} , with Copula-t distribution, a perfect tool for generating vectors with controlled correlation, with 4 degrees of freedom whose correlation matrix parameter is given by $\mathbf{R} = \mathbf{R}_t \otimes \mathbf{R}_s$ where \mathbf{R}_t (\mathbf{R}_s) is a Toeplitz matrix whose first row is $[1 \ \rho_t \ \dots \ \rho_t^{\tau-1}]$ ($[1 \ \rho_s \ \dots \ \rho_s^{p-1}]$). Each sample of \mathbf{d} is partitioned to produce source vectors, $\mathbf{d}(k) = [s(k\tau) \ s(k\tau+1) \ \dots \ s((k+1)\tau-1)]$. Therefore, we obtain the source vectors as samples of a wide-sense cyclostationary process whose correlation structure in time direction and space directions are governed by the parameters ρ_t and ρ_s , respectively.

In the simulations, we consider a scenario with 7 sources and 20 mixtures, an i.i.d. Gaussian convolutive mixing system with order 7 and an extractor of order 8. We set $\rho_s = 0.5$, $\rho_t = 0.5$ and $\tau = 5$. We note that the sources are non-stationary in this case (we will cover stationary sources in the digital communication sources scenario).

Figure 1 shows the output total Signal energy to total Interference+Noise energy (over all outputs) Ratio (SINR) obtained for the proposed BCA algorithm (\bar{J}_e) for various sample lengths under 45dB SNR. SINR performance of Minimum Mean Square Error (MMSE) filter of the same order, which uses full information about mixing system and source/noise statistics, is also shown to evaluate the relative success of the proposed approach. A comparison has also been made with a gradient maximization of the criterion (kurtosis) of [32] (KurtosisMax.) and **Alg.2** of [33] where we take $k_{max} = 50$ and $l_{max} = 20$. We have obtained these methods from [2], [34]. As we did not encounter any convolutive BSS algorithm

with correlated source separation capability, we compared our algorithm with some well known convolutive ICA approaches.

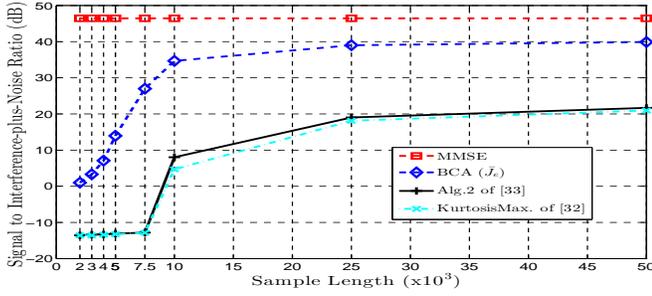


Fig. 1. Result of the proposed blind source extraction algorithm performance for the convolutive mixtures of dependent sources (ρ_s and ρ_t is set as 0.5) for various sample lengths under SNR = 45dB.

For the same setup, Figure 2 shows the output total Signal energy to total Interference+Noise energy (over all outputs) Ratio (SINR) obtained for the proposed BCA algorithm (\bar{J}_e), gradient maximization of the criterion (kurtosis) of [32] (KurtosisMax.), **Alg.2** of [33], and MMSE for various sample lengths under 20dB SNR.

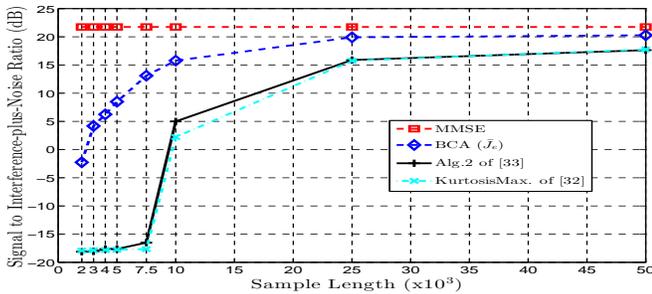


Fig. 2. Result of the proposed blind source extraction algorithm performance for the convolutive mixtures of dependent sources (ρ_s and ρ_t is set as 0.5) for various sample lengths under SNR = 20dB.

These results demonstrate that the performance of the proposed blind source extraction algorithm is approaching fast to its MMSE counterpart as the sample length increases. On the other hand, the performance of gradient maximization of the criterion (kurtosis) of [32] (KurtosisMax.) and **Alg.2** of [33] is far from the performance of MMSE filter even when the sample length is increased (Figure 1) or they require more sample lengths to reach the same SINR performance (Figure 2) since in the correlated case, independence assumption simply fails. Therefore, we observe that the proposed BCA approach is capable of blind source extraction of convolutive mixtures of space-time correlated sources.

B. Blind Source Separation

We first consider a similar scenario as in the blind source extraction examples to illustrate the performance of the proposed blind source separation algorithms regarding the separability of convolutive mixtures of space-time correlated sources.

Here, we consider a scenario with 5 sources and 15 mixtures, an i.i.d. Gaussian convolutive mixing system with order 5 and a separator of order 6 where the sample size is 50000.

Figure 3 shows the output total Signal energy to total Interference+Noise energy (over all outputs) Ratio (SINR) obtained for the proposed BCA algorithms (\bar{J}_{s1} , $\bar{J}_{s2,1}$, $\bar{J}_{s2,2}$, $\bar{J}_{s2,\infty}$) for various space correlation parameters under 45dB SNR. The performances of MMSE, gradient maximization of the criterion (kurtosis) of [32] (KurtosisMax.) and **Alg.2** of [33] are also plotted for comparison. We note that the algorithm \bar{J}_{s1} yields better performance than the other algorithms.

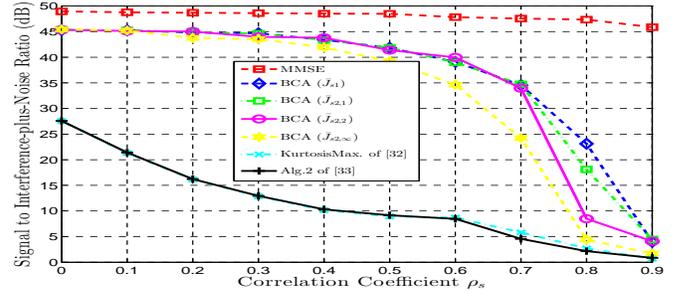


Fig. 3. Results of the proposed blind source separation algorithms' performances for the convolutive mixtures of dependent sources for various space correlation parameters under SNR = 45dB.

In Figure 4, we present a typical convergence graph of the BCA algorithms when $\rho_s = 0.3$ by plotting SINR performances and the objective functions versus the number of iterations.

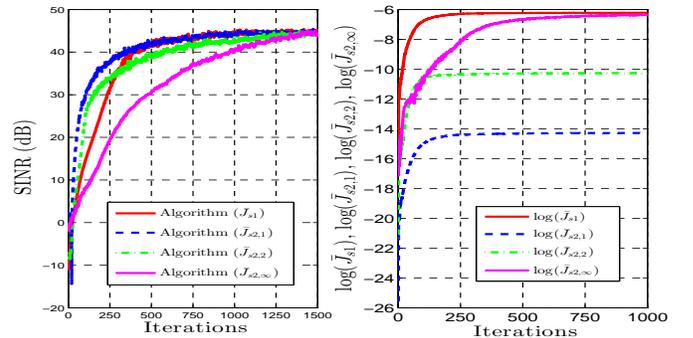


Fig. 4. Convergence graph of the introduced BCA algorithms.

For the same setup, Figure 5 shows the output total Signal energy to total Interference+Noise energy (over all outputs) Ratio (SINR) obtained for the BCA algorithm (\bar{J}_{s1}), gradient maximization of the criterion (kurtosis) of [32] (KurtosisMax.), **Alg.2** of [33], and MMSE for various space correlation parameters under 20dB SNR.

For the same setup, we finally choose 5dB SNR and Figure 6 shows the corresponding SINR results of the algorithms.

These results demonstrate that the performance of proposed blind source separation algorithms closely follow its MMSE counterpart for a wide range of correlation values. Therefore, we obtain a convolutive extension of the BCA approach introduced in [25], which is capable of separating convolutive mixtures of space-time correlated sources.

Also note that the proposed blind source separation algorithms maintain high separation performance for various space

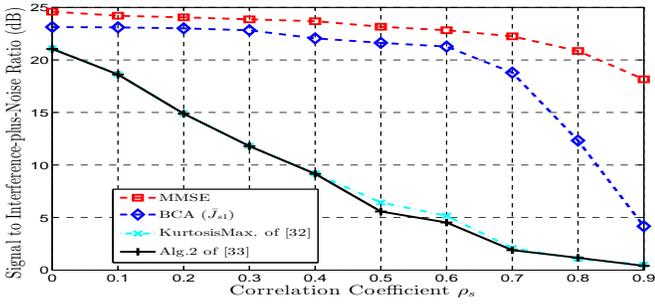


Fig. 5. Results of the proposed blind source separation algorithms' performances for the convolutive mixtures of dependent sources for various space correlation parameters under SNR = 20dB.

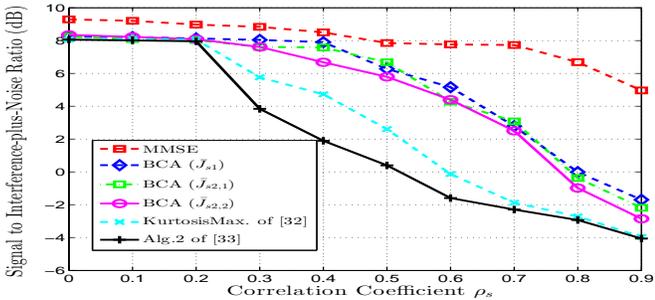


Fig. 6. Results of the proposed blind source separation algorithms' performances for the convolutive mixtures of dependent sources for various space correlation parameters under SNR = 5dB.

parameters. However, the performance of gradient maximization of the criterion (kurtosis) of [32] (KurtosisMax.) and **Alg.2** of [33] degrades substantially with increasing correlation since the independence assumption does not hold. We point out that when $\rho_s = 0$ the sources are independent, yet BCA algorithms still outperforms other ICA algorithms. This result can be attributed to the finite sample effects. In other words, although the sources are stochastically independent, finite samples may not reflect this behaviour and the sources may even have non-zero sample correlation. BCA algorithms being robust to such correlations can offer better performance. The effect of sample size is investigated in the next scenario.

We next consider the following scenario to illustrate the performance of the proposed blind source separation algorithm for the convolutive mixtures of digital communication sources. We consider 5 complex 4-QAM sources where we take 15 mixtures, an i.i.d. Gaussian convolutive mixing system with order 5 and a separator of order 6. The sources are stationary in this case. We use the objective function \bar{J}_{cs1} as the BCA algorithm for this simulation. The resulting Signal to Interference Ratio is plotted with respect to the sample lengths in Figure 7. We have also compared our algorithm with a gradient maximization of the criterion (kurtosis) of [32] (KurtosisMax.) and **Alg.2** of [33].

As it can be observed from Figure 7, the proposed BCA approach achieves better performance than ICA based approaches. We again note that, the proposed method does not assume/exploit statistical independence. The only impact of short data length is on accurate representation of source box

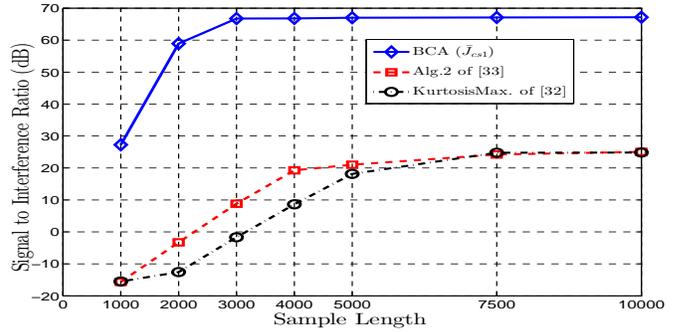


Fig. 7. Result of the proposed blind source separation algorithm performance for the convolutive mixtures of digital communication sources for various sample lengths.

boundaries. The simulation results suggest that the shorter data records may not be sufficient to reflect the stochastic independence of the sources, and therefore, the compared algorithms require more data samples to achieve the same SIR level as the proposed approach.

In the last scenario, the mixing system is chosen as (2 x 2) paraunitary with order 4 and the sources are 16-QAM. We use the objective function \bar{J}_{cs1} as the BCA algorithm for this simulation. The resulting Signal to Interference Ratio is plotted with respect to the sample lengths in Figure 8. We have also compared our algorithm with a gradient maximization of the criterion (kurtosis) of [32] (KurtosisMax.), **Alg.2** of [33], and the algorithm introduced in [35]¹.

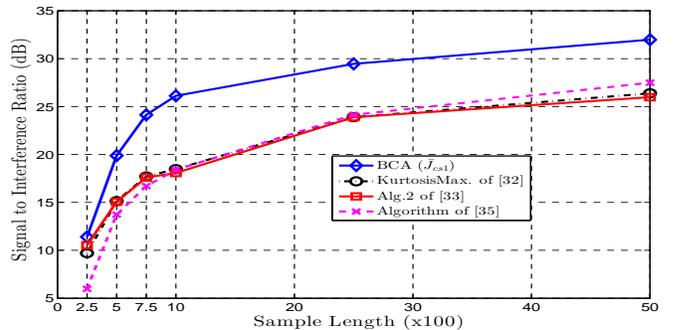


Fig. 8. Result of the proposed algorithm performance for the convolutive mixtures of digital communication sources by a paraunitary channel for various sample lengths.

We note that the proposed BCA approach achieves better performance. We also point out that the algorithm introduced in [35] assumes paraunitary channel and it requires numerically stable prewhitening operation for general equalizable channels [31] while the approach introduced in this paper does not assume any structure for the mixing system beyond equalizability.

VI. CONCLUSION

In this article, we introduced deterministic and geometric BCA frameworks for the convolutive blind source extraction

¹The authors would like to thank Prof. Lathauwer for kindly sharing their code for the algorithm in [35]

and separation problems. Contrary to the convolutive BCA framework in [26], [27] for stationary sources, the proposed deterministic framework is applicable to more general class of stationary and non-stationary sources. We should note that the proposed framework doesn't take advantage of the any special form of non-stationarity feature, rather it provides a flexible framework enabling separation of non-stationary signals in addition to stationary signals. Since the independence assumption is replaced by a much weaker domain separability assumption, the proposed framework is also flexible in terms of its ability to extract/separate dependent/correlated as well as independent sources. The numerical examples illustrate the dependent/independent source extraction/separation capability and potential short data record performance advantage of the proposed algorithms.

APPENDIX

A. Proof of Theorem 1

We first note that, following similar steps as in [25], when the assumption (A1) stated in Section II holds, we can write the range of o as $\tilde{\mathcal{R}}(o) = \|\tilde{\mathbf{g}}^T \tilde{\Lambda}\|_1$ where $\tilde{\Lambda} = I \otimes \Lambda$ is the range matrix of $\tilde{\mathbf{s}}_P$.

Since $o(l) = \tilde{\mathbf{g}}^T \tilde{\mathbf{s}}_P(l + M - 1)$ for $l = 1, 2, \dots, N_1$, we have

$$\begin{aligned} & \frac{1}{N_1} \sum_{l=1}^{N_1} (o(l) - \hat{\mu}_o)^2 \\ &= \frac{1}{N_1} \sum_{l=M}^N \tilde{\mathbf{g}}^T (\tilde{\mathbf{s}}_P(l) - \hat{\boldsymbol{\mu}}_{\tilde{\mathbf{s}}_P}) (\tilde{\mathbf{s}}_P(l) - \hat{\boldsymbol{\mu}}_{\tilde{\mathbf{s}}_P})^T \tilde{\mathbf{g}} \\ &= \frac{1}{N_1} \sum_{l=M}^N \tilde{\mathbf{g}}^T \tilde{\Lambda} (\tilde{\mathbf{s}}_P(l) - \hat{\boldsymbol{\mu}}_{\tilde{\mathbf{s}}_P}) (\tilde{\mathbf{s}}_P(l) - \hat{\boldsymbol{\mu}}_{\tilde{\mathbf{s}}_P})^T \tilde{\Lambda}^T \tilde{\mathbf{g}} \\ &= \tilde{\mathbf{g}}^T \tilde{\Lambda} \hat{\mathbf{R}}_{\tilde{\mathbf{s}}_P} \tilde{\Lambda}^T \tilde{\mathbf{g}}, \end{aligned} \quad (17)$$

where $\hat{\boldsymbol{\mu}}_{\tilde{\mathbf{s}}_P} = \frac{1}{N_1} \sum_{l=M}^N \tilde{\mathbf{s}}_P(l)$, $\hat{\boldsymbol{\mu}}_{\tilde{\mathbf{s}}_P} = \frac{1}{N_1} \sum_{l=M}^N \tilde{\mathbf{s}}_P(l)$ and $\hat{\mathbf{R}}_{\tilde{\mathbf{s}}_P}$ is defined as the sample covariance matrix of $\tilde{\mathbf{s}}_P$.

We can further define $\mathbf{q}^T = \tilde{\mathbf{g}}^T \tilde{\Lambda}$ and rewrite the equality (1) in terms of \mathbf{q} as

$$J_e(\mathbf{q}) = \frac{\sqrt{\mathbf{q}^T \hat{\mathbf{R}}_{\tilde{\mathbf{s}}_P} \mathbf{q}}}{\|\mathbf{q}\|_1}. \quad (18)$$

Note that, maximizing $J_e(\mathbf{q})$ is equivalent to the corresponding optimization setting

$$\begin{aligned} & \text{maximize} \quad \sqrt{\mathbf{q}^T \hat{\mathbf{R}}_{\tilde{\mathbf{s}}_P} \mathbf{q}} \\ & \text{s.t.} \quad \|\mathbf{q}\|_1 \leq \gamma \end{aligned}$$

where γ is a constant. Also note that, assuming $\hat{\mathbf{R}}_{\tilde{\mathbf{s}}_P} \succ 0$, $\sqrt{\mathbf{q}^T \hat{\mathbf{R}}_{\tilde{\mathbf{s}}_P} \mathbf{q}}$ is a convex function and the region of $\|\mathbf{q}\|_1 \leq \gamma$ corresponds to a convex polytope. From the definition of a convex polytope (Vertex Representation [36]), this is the convex hull of the vertices of polytope. Therefore, the maximum of $\sqrt{\mathbf{q}^T \hat{\mathbf{R}}_{\tilde{\mathbf{s}}_P} \mathbf{q}}$ will be attained at one of the vertices (whichever has the maximum value) and therefore, the maximum will be attained when \mathbf{q} has only one non-zero component. To see

that, we can take any vector \mathbf{q}_i inside the convex polytope (i.e., satisfying $\|\mathbf{q}_i\|_1 \leq \gamma$). From the definition of vertex representation [36], $\mathbf{q}_i = \alpha_1 \mathbf{q}_{v_1} + \alpha_2 \mathbf{q}_{v_2} + \dots + \alpha_{pP} \mathbf{q}_{v_{pP}}$ where $\mathbf{q}_{v_1}, \mathbf{q}_{v_2}, \dots, \mathbf{q}_{v_{pP}}$ are vertices of polytope and $\sum_{l=1}^{pP} \alpha_l = 1$. Defining $f(vq) = \sqrt{\mathbf{q}^T \hat{\mathbf{R}}_{\tilde{\mathbf{s}}_P} \mathbf{q}}$ and using Jensen's inequality, we have

$$\begin{aligned} f(\mathbf{q}) &\leq \alpha_1 f(\mathbf{q}_{v_1}) + \alpha_2 f(\mathbf{q}_{v_2}) + \dots + \alpha_{pP} f(\mathbf{q}_{v_{pP}}) \\ &\leq \max\{f(\mathbf{q}_{v_1}), f(\mathbf{q}_{v_2}), \dots, f(\mathbf{q}_{v_{pP}})\}. \end{aligned}$$

Therefore, the maximum is attained at the vertex which has the maximum value and this yields that \mathbf{q} has only one non-zero component.

To observe that from a geometric point of view, assuming $\hat{\mathbf{R}}_{\tilde{\mathbf{s}}_P} \succ 0$, for any constant γ , the vectors \mathbf{q} satisfying $\sqrt{\mathbf{q}^T \hat{\mathbf{R}}_{\tilde{\mathbf{s}}_P} \mathbf{q}} = \gamma$ constitutes an hyper-ellipsoid. Note that, for any constant value of $\|\mathbf{q}\|_1$, the maxima of $\sqrt{\mathbf{q}^T \hat{\mathbf{R}}_{\tilde{\mathbf{s}}_P} \mathbf{q}}$ will be attained at one of the corner points (i.e., where \mathbf{q} has only one non-zero component). A two dimensional example is illustrated in Figure 9.

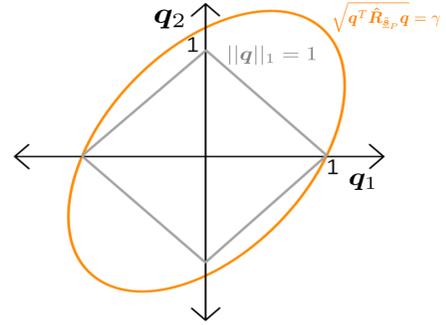


Fig. 9. Two dimensional example for the global maxima of (1).

Since $\tilde{\mathbf{g}}^T = \mathbf{q}^T \tilde{\Lambda}^{-1}$, $\tilde{\mathbf{g}}$ will also have only one non-zero component, therefore, the global maxima of (1) correspond to perfect extractors.

B. Proof of Theorem 2

We begin with noting that

$$\begin{aligned} \mathbb{R}\{o(k)\} &= \sum_{l=0}^{P-1} \mathbb{R}\{\mathbf{g}^T(l)\} \mathbb{R}\{\mathbf{s}(k + M - 1 - l)\} \\ &\quad - \mathbb{I}\{\mathbf{g}^T(l)\} \mathbb{I}\{\mathbf{s}(k + M - 1 - l)\}. \end{aligned}$$

If we define

$$\begin{aligned} \hat{\mathbf{g}} &= [\mathbb{R}\{\mathbf{g}^T(0)\} - \mathbb{I}\{\mathbf{g}^T(0)\} \dots - \mathbb{I}\{\mathbf{g}^T(P-1)\}]^T, \\ \hat{\mathbf{s}}_P(k) &= [\mathbb{R}\{\mathbf{s}^T(k)\} \mathbb{I}\{\mathbf{s}^T(k)\} \dots \mathbb{I}\{\mathbf{s}^T(k-P+1)\}]^T \end{aligned}$$

we then have

$$\mathbb{R}\{o(k)\} = \hat{\mathbf{g}}^T \hat{\mathbf{s}}_P(k + M - 1),$$

for $k = 1, 2, \dots, N_1$. Following similar steps, we can write the range of $\mathbb{R}\{o\}$ as $\tilde{\mathcal{R}}(\mathbb{R}\{o\}) = \|\hat{\mathbf{g}}^T \tilde{\Lambda}\|_1$ where $\tilde{\Lambda} = I \otimes \Lambda$

is the range matrix of \hat{s}_p . Similar to (17), we have

$$\frac{1}{N_1} \sum_{l=1}^{N_1} (\mathbb{R}\{o(l)\} - \mathbb{R}\{\hat{\mu}_o\})^2 = \hat{\mathbf{g}}^T \hat{\Lambda} \hat{\mathbf{R}}_{\hat{s}_p} \hat{\Lambda}^T \hat{\mathbf{g}},$$

where $\hat{\mathbf{R}}_{\hat{s}_p}$ is defined as the sample covariance matrix of \hat{s}_p . Defining $\hat{\mathbf{q}}^T = \hat{\mathbf{g}}^T \hat{\Lambda}$ and rewriting the equality (4) in terms of $\hat{\mathbf{q}}$ yields

$$J_{ce}(\hat{\mathbf{q}}) = \frac{\sqrt{\hat{\mathbf{q}}^T \hat{\mathbf{R}}_{\hat{s}_p} \hat{\mathbf{q}}}}{\|\hat{\mathbf{q}}\|_1}. \quad (19)$$

Following similar analogy, as a result, the maximum of (4) is attained when $\hat{\mathbf{g}}$ has only one non-zero component which also implies that $\hat{\mathbf{g}}$ has only one non-zero component. Note that the non-zero component of $\hat{\mathbf{g}}$ will be real or purely imaginary. Therefore, the global maxima of (4) correspond to a subset of perfect extractors for complex signals.

C. Proof of Theorem 3

We define the operator Γ_K such that $\Gamma_K(\tilde{\mathbf{G}})$ is a block Toeplitz matrix of dimension $Kp \times (K+P-1)p$ whose first block row is $[\mathbf{G}(0) \ \mathbf{G}(1) \ \dots \ \mathbf{G}(P-1) \ \mathbf{0} \ \dots \ \mathbf{0}]$ and first block column is $[\mathbf{G}^T(0) \ \mathbf{0} \ \dots \ \mathbf{0}]^T$ where the zero matrices ($\mathbf{0}$) have the size $p \times p$ same as the matrices $\mathbf{G}(l)$ for $l = 0, \dots, P-1$. This yields,

$$\tilde{o}_K(l) = \Gamma_K(\tilde{\mathbf{G}}) \tilde{s}_{K+P-1}(l+M-1),$$

for $l = K, K+1, \dots, N_1$. Defining $A = K+P-1$, we have

$$\hat{\mathbf{R}}_{\tilde{o}_K} = \Gamma_K(\tilde{\mathbf{G}}) \check{\Lambda} \hat{\mathbf{R}}_{\tilde{s}_A} \check{\Lambda}^T \Gamma_K(\tilde{\mathbf{G}})^T,$$

where $\check{\Lambda} = I \otimes \Lambda$ is the range matrix of \tilde{s}_A and $\hat{\mathbf{R}}_{\tilde{s}_A}$ is the sample covariance matrix of \tilde{s}_A . Defining $\mathbf{Q} = \Gamma_K(\tilde{\mathbf{G}}) \check{\Lambda}$ yields $\hat{\mathbf{R}}_{\tilde{o}_K} = \mathbf{Q} \hat{\mathbf{R}}_{\tilde{s}_A} \mathbf{Q}^T$.

Following similar steps as in [25], under the assumption (A1) stated in Section II, we can write the range of m^{th} component of \mathbf{o} as $\hat{\mathcal{R}}(o_m) = \|\tilde{\mathbf{G}}_{m,:} \check{\Lambda}\|_1$. Note that, $\|\tilde{\mathbf{G}}_{m,:} \check{\Lambda}\|_1 = \|\mathbf{Q}_{m,:}\|_1$ for $m = 1, 2, \dots, p$. Therefore, the range vector for the separator outputs can be rewritten as

$$\hat{\mathcal{R}}(\mathbf{o}) = [\|\mathbf{Q}_{1,:}\|_1 \ \|\mathbf{Q}_{2,:}\|_1 \ \dots \ \|\mathbf{Q}_{p,:}\|_1].$$

Rewriting the equality (7) in terms of \mathbf{Q} , we obtain

$$J_{s1}(\tilde{\mathbf{W}}) = \frac{\left(\sqrt{\det(\mathbf{Q} \hat{\mathbf{R}}_{\tilde{s}_A} \mathbf{Q}^T)}\right)^{1/K}}{\prod_{m=1}^p \|\mathbf{Q}_{m,:}\|_1}. \quad (20)$$

We note that for any $\tilde{\mathbf{G}}$ whose rows are not linearly independent we have $\det(\mathbf{Q} \hat{\mathbf{R}}_{\tilde{s}_A} \mathbf{Q}^T) = 0$, therefore, corresponding $\tilde{\mathbf{G}}$ can not be global maxima of (7). Hence for any $\tilde{\mathbf{G}}$ whose rows are linearly independent, assuming $\hat{\mathbf{R}}_{\tilde{s}_{K+P-1}} = \hat{\mathbf{R}}_{\tilde{s}_A} \succ 0$, to complete \mathbf{Q} into a full rank square matrix we introduce a $(P-1)p \times Ap$ matrix $\mathbf{M} = \mathbf{D}\mathbf{P}$ where $\mathbf{D} = \text{diag}(a_1, a_2, \dots, a_{(P-1)p})$ is a full rank diagonal matrix and \mathbf{P} is a permutation matrix such that $\det(\mathbf{M}\mathbf{B}\mathbf{M}^T) = 1$ where

we define $\mathbf{B} = \hat{\mathbf{R}}_{\tilde{s}_A} - \hat{\mathbf{R}}_{\tilde{s}_A} \mathbf{Q}^T (\mathbf{Q} \hat{\mathbf{R}}_{\tilde{s}_A} \mathbf{Q}^T)^{-1} \mathbf{Q} \hat{\mathbf{R}}_{\tilde{s}_A}$. This yields,

$$\begin{aligned} \det\left(\begin{bmatrix} \mathbf{Q} \\ \mathbf{M} \end{bmatrix} \hat{\mathbf{R}}_{\tilde{s}_A} \begin{bmatrix} \mathbf{Q}^T & \mathbf{M}^T \end{bmatrix}\right) &= \det(\mathbf{Q} \hat{\mathbf{R}}_{\tilde{s}_A} \mathbf{Q}^T) \\ \det\left(\mathbf{M} \left(\hat{\mathbf{R}}_{\tilde{s}_A} - \hat{\mathbf{R}}_{\tilde{s}_A} \mathbf{Q}^T (\mathbf{Q} \hat{\mathbf{R}}_{\tilde{s}_A} \mathbf{Q}^T)^{-1} \mathbf{Q} \hat{\mathbf{R}}_{\tilde{s}_A}\right) \mathbf{M}^T\right) & \\ = \det(\mathbf{Q} \hat{\mathbf{R}}_{\tilde{s}_A} \mathbf{Q}^T) \det(\mathbf{M}\mathbf{B}\mathbf{M}^T) &= \det(\mathbf{Q} \hat{\mathbf{R}}_{\tilde{s}_A} \mathbf{Q}^T). \end{aligned}$$

We note that $\begin{bmatrix} \mathbf{Q} \\ \mathbf{M} \end{bmatrix} \hat{\mathbf{R}}_{\tilde{s}_A} \begin{bmatrix} \mathbf{Q}^T & \mathbf{M}^T \end{bmatrix} \succ 0$ and $\mathbf{M}\mathbf{B}\mathbf{M}^T$ is the Schur complement of $\mathbf{Q} \hat{\mathbf{R}}_{\tilde{s}_A} \mathbf{Q}^T$, therefore, $\mathbf{M}\mathbf{B}\mathbf{M}^T \succ 0$. We also note that $\det(\mathbf{M}\mathbf{B}\mathbf{M}^T) = a_1^2 a_2^2 \dots a_{(P-1)p}^2 \det([\mathbf{B}]_{per})$ where $[\mathbf{B}]_{per}$ has the chosen rows and columns of \mathbf{B} depending on the positions of $a_1, a_2, \dots, a_{(P-1)p}$. Hence by choosing appropriate values for $a_1, a_2, \dots, a_{(P-1)p}$ we can obviously introduce a matrix \mathbf{M} such that

$$\det\left(\begin{bmatrix} \mathbf{Q} \\ \mathbf{M} \end{bmatrix} \hat{\mathbf{R}}_{\tilde{s}_A} \begin{bmatrix} \mathbf{Q}^T & \mathbf{M}^T \end{bmatrix}\right) = \det(\mathbf{Q} \hat{\mathbf{R}}_{\tilde{s}_A} \mathbf{Q}^T).$$

Using Hadamard's Inequality [37] yields

$$\begin{aligned} \det\left(\begin{bmatrix} \mathbf{Q} \\ \mathbf{M} \end{bmatrix} \hat{\mathbf{R}}_{\tilde{s}_A} \begin{bmatrix} \mathbf{Q}^T & \mathbf{M}^T \end{bmatrix}\right) \\ \leq \prod_{m=1}^{Kp} \|\mathbf{Q}_{m,:}\|_2^2 \prod_{n=1}^{(P-1)p} \|\mathbf{M}_{n,:}\|_2^2 \det(\hat{\mathbf{R}}_{\tilde{s}_A}). \end{aligned} \quad (21)$$

Note that $\prod_{m=1}^{Kp} \|\mathbf{Q}_{m,:}\|_2^2 = \left(\prod_{m=1}^p \|\mathbf{Q}_{m,:}\|_2^2\right)^K$ since \mathbf{Q} is block Toeplitz matrix. Hence,

$$\begin{aligned} \left(\sqrt{\det(\mathbf{Q} \hat{\mathbf{R}}_{\tilde{s}_A} \mathbf{Q}^T)}\right)^{1/K} &\leq \\ \left(\prod_{m=1}^p \|\mathbf{Q}_{m,:}\|_2\right) \left(\prod_{n=1}^{(P-1)p} \|\mathbf{M}_{n,:}\|_2\right)^{1/K} &\det(\hat{\mathbf{R}}_{\tilde{s}_A})^{1/2K}. \end{aligned}$$

Therefore, we have

$$\begin{aligned} J_{s1}(\tilde{\mathbf{W}}) &= \frac{\left(\sqrt{\det(\mathbf{Q} \hat{\mathbf{R}}_{\tilde{s}_A} \mathbf{Q}^T)}\right)^{1/K}}{\prod_{m=1}^p \|\mathbf{Q}_{m,:}\|_1} \\ &\leq \frac{\prod_{m=1}^p \|\mathbf{Q}_{m,:}\|_2}{\prod_{m=1}^p \|\mathbf{Q}_{m,:}\|_1} \left(\prod_{n=1}^{(P-1)p} \|\mathbf{M}_{n,:}\|_2\right)^{1/K} \det(\hat{\mathbf{R}}_{\tilde{s}_A})^{1/2K} \\ &\leq \left(\prod_{n=1}^{(P-1)p} \|\mathbf{M}_{n,:}\|_2\right)^{1/K} \det(\hat{\mathbf{R}}_{\tilde{s}_A})^{1/2K}, \end{aligned} \quad (22)$$

due to the ordering $\|\mathbf{q}\|_1 \geq \|\mathbf{q}\|_2$ for any \mathbf{q} .

To achieve the equality in (22), the equalities $\|\mathbf{Q}_{m,:}\|_1 = \|\mathbf{Q}_{m,:}\|_2$ for $m = 1, 2, \dots, p$ and the equality in (21) should be achieved. The equalities $\|\mathbf{Q}_{m,:}\|_1 = \|\mathbf{Q}_{m,:}\|_2$ for $m = 1, 2, \dots, p$ are achieved if and only if the first p rows of \mathbf{Q} has only one non-zero element. Since $\Gamma_K(\tilde{\mathbf{G}}) = \mathbf{Q} \check{\Lambda}^{-1}$, this implies that each row of $\tilde{\mathbf{G}}$ has only one non-zero element. The inequality in (21) is achieved if and only if the rows of \mathbf{Q} are perpendicular to each other and to the rows of \mathbf{M} which

yields that the rows of $\Gamma_K(\tilde{\mathbf{G}})$ are perpendicular to each other and to the rows of \mathbf{M} . Note that since $K \geq P$, the structure of $\Gamma_K(\tilde{\mathbf{G}})$ guarantees that there is a block column which contains $\mathbf{G}(0), \mathbf{G}(1), \dots, \mathbf{G}(P-1)$, therefore, the non-zero entries of $\tilde{\mathbf{G}}$ would not be in the same position with respect to mod p , since otherwise $J_{s1}(\tilde{\mathbf{W}})$ would be simply 0.

As a result, the maximum is achieved if and only if $\tilde{\mathbf{G}}$ corresponds to perfect separator transfer matrix in the form $\mathbb{G}(z) = \text{diag}(\alpha_1 z^{-d_1}, \alpha_2 z^{-d_2}, \dots, \alpha_p z^{-d_p}) \mathbf{P}$ where $\mathbb{G}(z)$ is the Z-transform of the overall system function $\{\mathbf{G}(l); l \in \{0, \dots, P-1\}\}$, α_k 's are non-zero real scalings, and d_k 's are non-negative integer delays.

We note that due to the structure of the set of global maxima of the objective function $J_{s1}(\tilde{\mathbf{W}})$, the sources are recovered up to scaling and order. These indeterminacy cannot be resolved without any extra information in BSS problems.

Here, we point out that the blind source extraction problem is a special case of the blind source separation problem. Therefore, this proof can simply be also applied to the blind source extraction method. However, we treat the blind source extraction problem as a separate case to provide alternative geometric intuition.

D. Analysis of the Family of Objective Functions ($J_{s2,r}$)

Before analysing this family of objective functions for some special r values, similar to the proof of Theorem 3, we can rewrite (8) in terms of \mathbf{Q} and obtain

$$J_{s2,r}(\tilde{\mathbf{W}}) = \frac{\left(\sqrt{\det(\mathbf{Q}\hat{\mathbf{R}}_{\underline{\mathbf{s}}_A}\mathbf{Q}^T)}\right)^{1/K}}{\left\|\left[\|\mathbf{Q}_{1,:}\|_1 \quad \|\mathbf{Q}_{2,:}\|_1 \quad \dots \quad \|\mathbf{Q}_{p,:}\|_1\right]^T\right\|_r^p}.$$

Following similar steps, by modifying (22), we can obtain the corresponding inequality

$$J_{s2,r}(\tilde{\mathbf{W}}) \leq \frac{\prod_{m=1}^p \|\mathbf{Q}_{m,:}\|_2}{\left\|\left[\|\mathbf{Q}_{1,:}\|_1 \quad \|\mathbf{Q}_{2,:}\|_1 \quad \dots \quad \|\mathbf{Q}_{p,:}\|_1\right]^T\right\|_r^p} \left(\prod_{n=1}^{(P-1)p} \|\mathbf{M}_{n,:}\|_2\right)^{1/K} \det(\hat{\mathbf{R}}_{\underline{\mathbf{s}}_A})^{1/2K}.$$

The results of analysing this family of objective functions, for some special r values:

- $r = 1$ Case: In this case, we have

$$\begin{aligned} & \left\|\left[\|\mathbf{Q}_{1,:}\|_1 \quad \|\mathbf{Q}_{2,:}\|_1 \quad \dots \quad \|\mathbf{Q}_{p,:}\|_1\right]^T\right\|_1^p \\ &= \left(\sum_{m=1}^p \|\mathbf{Q}_{m,:}\|_1\right)^p \geq p^p \prod_{m=1}^p \|\mathbf{Q}_{m,:}\|_1, \end{aligned}$$

where the inequality comes from Arithmetic-Geometric-Mean-Inequality, and the equality is achieved if and only if all the rows \mathbf{Q} have the same 1-norm. Hence, we have

$$\begin{aligned} J_{s2,1}(\tilde{\mathbf{W}}) &\leq \frac{\prod_{m=1}^p \|\mathbf{Q}_{m,:}\|_2}{p^p \prod_{m=1}^p \|\mathbf{Q}_{m,:}\|_1} \left(\prod_{n=1}^{(P-1)p} \|\mathbf{M}_{n,:}\|_2\right)^{\frac{1}{K}} \\ \det(\hat{\mathbf{R}}_{\underline{\mathbf{s}}_A})^{\frac{1}{2K}} &\leq \frac{1}{p^p} \left(\prod_{n=1}^{(P-1)p} \|\mathbf{M}_{n,:}\|_2\right)^{\frac{1}{K}} \det(\hat{\mathbf{R}}_{\underline{\mathbf{s}}_A})^{\frac{1}{2K}}. \end{aligned}$$

As a result, \mathbf{Q} is a global maximum of $J_{s2,1}(\tilde{\mathbf{W}})$ if and only if it is a perfect separator matrix of the form

$$\mathbf{Q} = k\mathbf{P}\text{diag}(\rho),$$

where k is a non-zero value, $\rho \in \{-1, 1\}^p$ and \mathbf{P} is a permutation matrix. This implies $\tilde{\mathbf{G}}$ is a global maximum of $J_{s2,1}(\tilde{\mathbf{W}})$ if and only if the corresponding form is satisfied

$$\Gamma_K(\tilde{\mathbf{G}}) = k\mathbf{P}\check{\Lambda}^{-1}\text{diag}(\rho).$$

Therefore, the global maxima of the objective function $J_{s2,1}$ corresponds to a subset of perfect separators.

- $r = 2$ Case: In this case, using the basic norm inequality and Arithmetic-Geometric-Mean-Inequality, for any $\mathbf{x} \in \mathbb{R}^p$, we have

$$\left(\|\mathbf{x}\|_2\right)^p \geq \left(\frac{1}{\sqrt{p}}\|\mathbf{x}\|_1\right)^p \geq p^{p/2} \prod_{m=1}^p |x_m|$$

where the equality is achieved if and only if all the components of \mathbf{x} are equal in magnitude. As a result, this yields

$$J_{s2,2}(\tilde{\mathbf{W}}) \leq \frac{1}{p^{p/2}} \left(\prod_{n=1}^{(P-1)p} \|\mathbf{M}_{n,:}\|_2\right)^{\frac{1}{K}} \det(\hat{\mathbf{R}}_{\underline{\mathbf{s}}_A})^{\frac{1}{2K}}.$$

Similarly, $J_{s2,2}$ has the same set of global maxima as $J_{s2,1}$.

- $r = \infty$ Case: Following similar steps, using the basic norm inequality and Arithmetic-Geometric-Mean-Inequality, for any $\mathbf{x} \in \mathbb{R}^p$, we have

$$\left(\|\mathbf{x}\|_\infty\right)^p \geq \left(\frac{1}{p}\|\mathbf{x}\|_1\right)^p \geq \prod_{m=1}^p |x_m|,$$

where the equality is achieved if and only if all the components of \mathbf{x} are equal in magnitude. Based on this inequality, we obtain

$$J_{s2,\infty}(\tilde{\mathbf{W}}) \leq \left(\prod_{n=1}^{(P-1)p} \|\mathbf{M}_{n,:}\|_2\right)^{\frac{1}{2K}} \det(\hat{\mathbf{R}}_{\underline{\mathbf{s}}_A})^{\frac{1}{2K}}.$$

Therefore, $J_{s2,\infty}$ also has same set of global optima as $J_{s2,1}$ and $J_{s2,2}$.

E. Proof of Theorem 4

We begin with observing that

$$\begin{aligned} \Re\{\mathbf{o}(k)\} &= \sum_{l=0}^{P-1} \Re\{\mathbf{G}^T(l)\} \Re\{\mathbf{s}(k+M-1-l)\} \\ &\quad - \Im\{\mathbf{G}^T(l)\} \Im\{\mathbf{s}(k+M-1-l)\}, \\ \Im\{\mathbf{o}(k)\} &= \sum_{l=0}^{P-1} \Im\{\mathbf{G}^T(l)\} \Re\{\mathbf{s}(k+M-1-l)\} \\ &\quad + \Re\{\mathbf{G}^T(l)\} \Im\{\mathbf{s}(k+M-1-l)\}. \end{aligned}$$

Defining $\hat{\mathbf{G}} =$

$$\begin{bmatrix} \Re\{\mathbf{G}_0\} & -\Im\{\mathbf{G}_0\} & \dots & \Re\{\mathbf{G}_{P-1}\} & -\Im\{\mathbf{G}_{P-1}\} \\ \Im\{\mathbf{G}_0\} & \Re\{\mathbf{G}_0\} & \dots & \Im\{\mathbf{G}_{P-1}\} & \Re\{\mathbf{G}_{P-1}\} \end{bmatrix} \quad \text{and}$$

$\hat{\mathbf{s}}_{K+P-1}(k) = [\mathbb{R}\{\mathbf{s}^T(k)\} \quad \mathbb{I}\{\mathbf{s}^T(k)\} \quad \dots \quad \mathbb{R}\{\mathbf{s}^T(k-K-P+2)\} \quad \mathbb{I}\{\mathbf{s}^T(k-K-P+2)\}]^T$ yields $\hat{\mathbf{d}}_K(k) = \Gamma_{2K}(\hat{\mathbf{G}})\hat{\mathbf{s}}_{K+P-1}(k)$. Thus,

$$\hat{\mathbf{R}}_{\hat{\mathbf{d}}_K} = \Gamma_{2K}(\hat{\mathbf{G}})\hat{\Lambda}\hat{\mathbf{R}}_{\hat{\mathbf{s}}_{K+P-1}}\hat{\Lambda}^T\Gamma_{2K}(\hat{\mathbf{G}})^T,$$

where $\hat{\Lambda} = I \otimes \Lambda$ is the range matrix of $\hat{\mathbf{s}}_{K+P-1}$ and $\hat{\mathbf{R}}_{\hat{\mathbf{s}}_{K+P-1}}$ is defined as the sample covariance matrix of $\hat{\mathbf{s}}_{K+P-1}$.

Defining $\hat{\mathbf{Q}} = \Gamma_{2K}(\hat{\mathbf{G}})\hat{\Lambda}$ and following similar steps, we can write $\prod_{m=1}^{2p} \hat{\mathcal{R}}(\hat{\mathbf{d}}_m) = \prod_{m=1}^{2p} \|\hat{\mathbf{Q}}_{m,:}\|_1$. Rewriting (12) in terms of $\hat{\mathbf{Q}}$ yields

$$J_{cs1}(\tilde{\mathbf{W}}) = \frac{\left(\sqrt{\det(\hat{\mathbf{Q}}\hat{\mathbf{R}}_{\hat{\mathbf{s}}_A}\hat{\mathbf{Q}}^T)}\right)^{1/K}}{\prod_{m=1}^{2p} \|\hat{\mathbf{Q}}_{m,:}\|_1}.$$

Note that we have the similar expression as (20). Hence, the proof of Theorem 3 also applies here. Note that the structure of $\Gamma_{2K}(\hat{\mathbf{G}})$ implies that the non-zero entries of $\hat{\mathbf{G}}$ can only be real or purely imaginary. Therefore, the set of global maxima for the objective function (12) corresponds to a subset of complex perfect separators.

REFERENCES

- [1] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*. John Wiley and Sons Inc., 2001.
- [2] P. Comon and C. Jutten, *Handbook of Blind Source Separation: Independent Component Analysis and Applications*. Academic Press, 2010.
- [3] P. Comon, "Independent component analysis, A new concept?" *Signal Processing*, vol. 36, pp. 287–314, April 1994.
- [4] A. Belouchrani, K. Abed-Meraim, J.-F. Cardoso, and E. Moulines, "A blind source separation technique using second-order statistics," *Signal Processing, IEEE Transactions on*, vol. 45, no. 2, pp. 434–444, Feb 1997.
- [5] M. Matsuoka, M. Ohaya, and M. Kawamoto, "A neural net for blind separation of nonstationary signals," *Neural Networks*, vol. 8, no. 3, pp. 411–419, March 1995.
- [6] P. Georgiev, F. Theis, and A. Cichocki, "Sparse component analysis and blind source separation of underdetermined mixtures," *IEEE Transactions on Neural Networks*, vol. 16, no. 4, pp. 992–996, July 2005.
- [7] A. J. van der Veen and A. Paulraj, "An analytical constant modulus algorithm," *IEEE Transactions on Signal Processing*, vol. 44, no. 5, pp. 1136–1155, May 1996.
- [8] P. Jallon, A. Chevreuril, and P. Loubaton, "Separation of digital communication mixtures with the CMA: Case of unknown symbol rates," *Signal Processing*, vol. 90, no. 9, pp. 2633–2647, September 2010.
- [9] S. Talwar, M. Viberg, and A. Paulraj, "Blind separation of synchronous co-channel digital signals using an antenna array. I. algorithms," *IEEE Transactions on Signal Processing*, vol. 44, no. 5, pp. 1184–1197, May 1996.
- [10] C. G. Puntonet and A. Prieto, "Neural net approach for blind separation of sources based on geometric properties," *Neurocomputing*, no. 18, pp. 141164, 1998.
- [11] D.-T. Pham, "Blind separation of instantaneous mixtures of sources based on order statistics," *IEEE Trans. on Signal Processing*, vol. 48, pp. 363–375, February 2000.
- [12] S. Cruces and I. Duran, "The minimum support criterion for blind source extraction: A limiting case of the strengthened young's inequality," *Lecture Notes in Computer Science*, pp. 57–64, September 2004.
- [13] F. Vrins, M. Verleysen, and C. Jutten, "SWM: A class of convex contrasts for source separation," *Proceedings of IEEE ICASSP 2005*, vol. V, pp. 161–164, March 2005.
- [14] F. Vrins, J. A. Lee, and M. Verleysen, "A minimum-range approach to blind extraction of bounded sources," *IEEE Trans. on Neural Networks*, vol. 18, no. 3, pp. 809–822, May 2007.
- [15] F. Vrins and D. Pham, "Minimum range approach to blind partial simultaneous separation of bounded sources: Contrast and discriminability properties," *Neurocomputing*, vol. 70, pp. 1207–1214, March 2007.
- [16] S. Vembu, S. Verdu, R. Kennedy, and W. Sethares, "Convex cost functions in blind equalization," *IEEE Trans. on Signal Processing*, vol. 42, pp. 1952–1960, August 1994.
- [17] A. T. Erdogan and C. Kizilkale, "Fast and low complexity blind equalization via subgradient projections," *IEEE Trans. on Signal Processing*, vol. 53, pp. 2513–2524, July 2005.
- [18] A. T. Erdogan, "A blind separation approach for magnitude bounded sources," *Proceedings of IEEE ICASSP 2005, Philadelphia*, vol. V, pp. 177–180, March 2005.
- [19] A. T. Erdogan, "A simple geometric blind source separation method for bounded magnitude sources," *IEEE Trans. on Signal Processing*, vol. 54, pp. 438–449, February 2006.
- [20] A. T. Erdogan, "Globally convergent deflationary instantaneous blind source separation algorithm for digital communication signals," *IEEE Trans. on Signal Processing*, vol. 55, no. 5, pp. 2182–2192, May 2007.
- [21] A. T. Erdogan, "Adaptive algorithm for the blind separation of sources with finite support," *Proceedings of EUSIPCO 2008, Lausanne*, August 2008.
- [22] S. Cruces, "Bounded component analysis of linear mixtures: A criterion for minimum convex perimeter," *IEEE Trans. on Signal Process.*, vol. 58, no. 4, pp. 2141–2154, April 2010.
- [23] P. Aguilera, S. Cruces, I. Duran, A. Sarmiento, and D. Mandic, "Blind separation of dependent sources with a bounded component analysis deflationary algorithm," *IEEE Signal Processing Letters*, vol. 20, no. 7, pp. 709–712, July 2013.
- [24] A. T. Erdogan, "On the convergence of symmetrically orthogonalized bounded component analysis algorithms for uncorrelated source separation," *IEEE Transactions on Signal Processing*, pp. 6058–6063, November 2012.
- [25] A. T. Erdogan, "A class of bounded component analysis algorithms for the separation of both independent and dependent sources," *IEEE Transactions on Signal Processing*, vol. 61, no. 22, pp. 5730–5743, November 2013.
- [26] H. A. Inan and A. T. Erdogan, "A bounded component analysis approach for the separation of convolutive mixtures of dependent and independent sources," *IEEE International Conference on Acoustics, Speech and Signal Processing*, May 2013.
- [27] H. A. Inan and A. T. Erdogan, "Convolutive Bounded Component Analysis Algorithms for Independent and Dependent Source Separation," *IEEE Transactions on Neural Networks and Learning Systems*, to appear.
- [28] M. D. Craig, "Minimum-volume transforms for remotely sensed data," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 32, no. 3, pp. 542–552, May 1994.
- [29] L. Miao and H. Qi, "Endmember Extraction From Highly Mixed Data Using Minimum Volume Constrained Nonnegative Matrix Factorization," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 45, no. 3, pp. 765–777, March 2007.
- [30] W. S. B. Ouedraogo, A. Souloumiac, M. Jaidane, and C. Jutten, "Simplex Cone Shrinking Algorithm for Unmixing Nonnegative Sources," *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on*, pp. 2405–2408, March 2012.
- [31] Y. Inouye and R.-W. 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.
- [32] C. Simon, P. Loubaton, and C. Jutten, "Separation of a class of convolutive mixtures: a contrast function approach," *Signal Processing*, vol. 81, no. 4, pp. 883–887, 2001.
- [33] M. Castella and E. Moreau, "New kurtosis optimization schemes for MISO equalization," *IEEE Transactions on Signal Processing*, vol. 60, no. 3, pp. 1319–1330, 2012.
- [34] P. Jallon, M. Castella, and R. Dubroca. (2011) Matlab toolbox for separation of convolutive mixtures. Available: <http://www-public.int-evry.fr/~castella/toolbox/>
- [35] L. De Lathauwer, "Algebraic Techniques for the Blind Deconvolution of Constant Modulus Signals," *Proc. 12th European Signal Processing Conference (EUSIPCO 2004)*, pp. 225–228, September 2004.
- [36] B. Grunbaum, V. Klee, A. Perles, and G. Shephard, *Convex Polytopes*. John Wiley & Sons, Incorporated, 1967.
- [37] D. Garling, *Inequalities: a journey into linear analysis*. Cambridge University Press, 2007.