Convergence Analysis for a Class of Source Separation Methods

Alper T. Erdogan
Electrical and Electronics Engineering
Koc University
Istanbul, Turkey
Email: alperdogan@ku.edu.tr

***Abstract*—Blind Source Separation (BSS) is a topic of interest in different fields of information processing with a large span of applications, e.g., in communications, pattern recognition, brain activity monitoring and audio processing . This article reviews some convergence analysis results for parallel BSS algorithms that are used for the simultaneous extraction of sources. The emphasis will be on two major BSS schemes, namely, Indepenent Component Analysis and Bounded Component Analysis. The article underlines the major convergence results regarding these two branches and their connections.**

***Index Terms*—Blind Source Separation, Independent Component Analysis, Bounded Component Analysis, Convergence Analysis**

I. Introduction

Blind Source Separation refers to the unsupervised separation of sources from their mixture samples, which has been a problem of interest in many different areas of Signal Processing and Machine Learning [1]. In fact, it is the central problem for various unrelated applications including Multiple-Input-Multiple-Output Communication Systems, Microphone Array Sound Separation, Brain Activity Monitoring, Financial Factor Analysis and many more. The richness and diverseness of the application pool for the BSS is due to its extreme flexibility provided by its unsupervised nature which eliminates the need for the training information. The power of this property is highlighted especially for applications where the use of training information is either too expensive or impossible.

Although the unsupervised nature results in a broad application span, it is also the main reason for the difficulty in solving this problem. The unavailability of the training data combined with the lack of mathematical description for the mixing system makes it challenging to come up with adaptive algorithms capable of separating original sources using only their mixture samples. This difficulty has been overcome by the use of some additional side information and/or assumptions regarding the data model. In fact, we can classify the existing BSS approaches based on the underlying assumptions they make:

- *Statistical Independence:* The most common assumption is the mutual statistical independence of sources. The approach based on this assumption is referred as Independent Component Analysis (ICA) and it is the most widespread BSS approach. In fact, sometimes it is, mistakenly, considered synonymous to BSS. As the main solution recipe for the BSS problem, ICA emerged before the first half of 90s, see [2] and the references therein. It can be, alternatively perceived as the extension of the Principal Component Analysis approach, which decomposes a vector as sum of some uncorrelated components. ICA extends the partial, second order moment based, statistical decoupling in PCA to the complete statistical decoupling.
The basic principal of all existing ICA methods is achieving mutual independence of the separator outputs. There are different approaches in achieving this major goal which are inherently connected with each other. One approach is based on the relationship between the level of output independence and the level of output nongaussianity. According to this approach maximizing nongaussianity is equivalent to maximizing independence. For the measurement of nongaussianity typical, choices are the higher order statistic based Kurtosis [3] and source pdf based negentropy [4] objective functions.
Another approach is the minimization of output mutual information, see for example [5]. The use of Maximum Likelihood estimation method in ICA framework [6] is another alternative but related approach. The list of the proposed ICA approaches and the relevant algorithmic work is a rather long list. However, FastICA [7] stands out as probably the most popular ICA, which successfully uses contrast function technique connected to the aforementioned ICA approaches in conjunction with the fixed point algorithm implementation.
- *Time-Sample Structure:* In the ICA framework, the sources are assumed to have independent identically distributed (i.i.d.) temporal structure. The corresponding algorithms rely on the explicit or implicit use of Higher Order (than second order) Statistics (HOS). The use of HOS typically is reflected as complex algorithms that demand for relatively large number of mixtures for reasonable separation performance. Contrary to these, there are BSS approaches that exploit potential time structure in data such as correlation such as [8] , for which the Second Order Statistics (SOS) is sufficient. Similarly, it was shown that the temporal distribution variation in sources, i.e. nonstationarity, can be exploited to achieve separation through a decorrelation procedure

978-1-4577-0746-9/11/$26.00 ©2011 IEEE

[9]. Another example of time structure used for separation is the cyclostationary property of digital communication sources, see for example [10] and the references therein.

- *Sparseness:* Sparseness is another structure exploited in BSS leading to Sparse Component Analysis (SCA) [11] The sparsity of the speech sources in time-frequency plane can also be exploited for the audio separation purposes [12].
- *Special Distribution Structure of Communication Signals:* The extension of the blind equalization of single communication channels led to the special BSS algorithms targeted for MIMO communication applications. Among these we can list the Constant Modulus Algorithm, utilizing constant magnitude structure of FM and QPSK communication signals [13], [14] and the Finite Alphabet Algorithm utilizing special constellation structure of digital communication signals [15].
- *Boundedness:* In typical practical BSS applications source values take their values from a compact set. This property has been exploited especially in some recent ICA algorithms. The potential for utilizing boundedness property in ICA framework was first put forward in [16]. In this work, Pham reformulated the mutual information cost function in terms of order statistics. In the bounded case, this formulation leads to the effective minimization of the separator output ranges. The reference [17] introduced output infinity norm minimization based geometric approach for the separation of bounded sources with peak symmetry, as an extension of convex optimization based blind equalization algorithms [18], [19]. In this article, the equivalence of global optima of the corresponding cost function and the perfect separators is proven. In addition, a subgradient search based adaptive algorithm with low complexity update rule is provided. The same article also illustrated that the exploitation of the boundedness property resulted in superior performance compared to conventional ICA schemes such as FastICA [7] and Multi-user Kurtosis [20]. This performance advantage is magnified especially for short data bursts. The proposed BSS approach was employed for enhancing the data rates of optical-fiber communication systems [21], [22]. In reference [23], a globally convergent algorithm is provided for the same bounded-magnitude BSS framework. Around the same time frame, Vrins et. al. independently contributed to the area of ICA for sources with finite support [24], [25].

In a recent article [26], it was shown that under the generic boundedness assumption about the sources, a more general framework than ICA can be constructed. This framework is named Bounded Component Analysis (BCA) and it makes weaker assumptions about the relation among the sources. According to BCA approach, if the sources are bounded, the independence assumption can be replaced with a more general assumption about domain separability, which can be stated as "(the convex hull of the) support set of the joint distribution of sources can be written as the Cartesian product of (the convex hull of the) support sets of the individual source distributions. Note that independence assumption implies domain separability condition, however, the reverse is not necessarily true. Therefore under the source boundedness assumption BCA forms a more general framework which allows separation of dependent sources and which includes ICA as a special case.

In this article, we review some results regarding the convergence of both ICA and BCA algorithms. The focus will be on the BSS algorithms extracting sources in parallel (instead of sequential peeling approach). For this purpose, in Section II we introduce the setup for the assumed BSS problem. In Section III, we discuss parallel ICA and BCA algorithms. The discussion on the comparative convergence analysis results is provided in Section IV. Finally, Section V is the conclusion.

II. Blind Source Separation Setup

The basic setup for the BSS problem is shown in Figure 1: In this set up

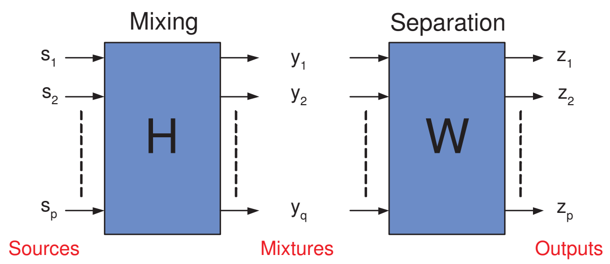

Fig. 1. Simulation Results

- $s_1(k), s_2(k), \ldots s_p(k)$ area the desired sources. Without loss of generality, they are assumed to be zero mean signals. We also define the source vector sequence as

$$\mathbf{s}(k) = \begin{bmatrix} s_1(k) \\ s_2(k) \\ \vdots \\ s_p(k) \end{bmatrix}. \tag{1}$$

- The memoryless mixing system is represented with the matrix $\mathbf{H} \in \Re^{q \times p}$, where q, the number of mixtures is assumed to be greater than or equal to p, the number of sources, i.e., the (over)determined case. For the invertibility of the mixing, we further assume that $\mathbf{H}$ is full-rank.
- $y_1(k), y_2(k), \ldots, y_q(k)$ are the mixture sequences. We define the mixture vector sequence as

$$\mathbf{y}(k) = \begin{bmatrix} y_1(k) \\ y_2(k) \\ \vdots \\ y_q(k) \end{bmatrix}. \tag{2}$$

The relation between the mixture vectors and source vectors is simply given by

$$\mathbf{y}(k) = \mathbf{H}\mathbf{x}(k). \tag{3}$$

- The separation system is represented by the matrix $\mathbf{W} \in \Re^{p \times q}$.
- $z_1(k), z_2(k), \ldots, z_p(k)$ are the separator outputs. The separator output vector sequence is defined as

$$\mathbf{z}(k) = \begin{bmatrix} z_1(k) \\ z_2(k) \\ \vdots \\ z_p(k) \end{bmatrix}. \tag{4}$$

Based on these definitions, the separator output vector can be written as

$$\mathbf{z}(k) = \mathbf{W}\mathbf{y}(k). \tag{5}$$

- We also define the overall mapping from sources to separator outputs as the matrix $\mathbf{G} \in \Re^{p \times p}$, which is related to mixing and separation processes by

$$\mathbf{G} = \mathbf{W}\mathbf{H}. \tag{6}$$

- The goal of the BSS problem is to train the separator matrix $\mathbf{W}$ using the available set of mixture samples $\mathcal{Y} = \{\mathbf{y}(1), \mathbf{y}(2), \ldots, \mathbf{y}(L)\}$ such that the separator outputs are as close as possible to the original sources. We should note that there are two fundamental uncertainties that can not be overcome unless some additional side information exists. These are
  - *Permutation Ambiguity*: It is not possible to determine the exact ordering of the sources,
  - *Sign Ambiguity*: It is not possible to determine the sign of the individual sources.

  Therefore, under the existence of these ambiguities, the goal of expecting $\mathbf{W}$ to be equal to a left inverse of $\mathbf{H}$ is too ambitious. It is only reasonable to expect $\mathbf{W}$ to converge to vicinity of the so-called perfect separator matrices for which the overall mapping satisfies the relation

$$\mathbf{G} = \mathbf{D}\mathbf{P}. \tag{7}$$

Here $\mathbf{P}$ is a permutation matrix, corresponding to the permutation ambiguity, and $\mathbf{D}$ is a diagonal matrix with $\{\pm 1\}$ entries at the diagonal, corresponding to sign ambiguity.

III. PARALLEL BSS ALGORITHMS

In order to achieve the goal of separation outlined in the previous section, there are two fundamental approaches:

- *Sequential Separation Methods:* In this approach, the sources are extracted from the mixture sequentially. In this so-called Deflationary scheme, sources are extracted one at a time. Once a source is obtained, its contribution to the mixture is eliminated and the procedure is repeated for the remaining sources. This onion peeling approach is convenient due to the relative simplicity of the single source extraction process. In addition the convergence analysis in this case is more tractable (see for example [23]) compared to its parallel counterpart. However, the error propagation is the main practical issue concerning deflationary algorithms. The errors made in the extraction of the previous sources results in error growth in the following stages.
- *Parallel Separation Methods:* The class of BSS algorithms under this category extracts all sources in parallel. It is essentially a vector extraction process where the size of the vector is equal to the number of sources. Although performance-wise this is a better approach, the convergence analysis in this case is more challenging.

The focus of this article is to highlight convergence results for the algorithms in the second class, namely parallel BSS algorithms for both ICA and BCA schemes. The parallel ICA and BCA algorithms are briefly discussed in the following subsections.

A. Parallel ICA Algorithms

As noted in the introduction section, the ICA algorithms exploit source independence to train the separator matrix. The goal of achieving independence at the separator output is equivalent to the goal of obtaining "perfect separators" defined in the previous subsection.

The goal of achieving independence is typically implemented by a two step procedure:

1. *Whitening:* In the first step, first the uncorrelated mixtures are obtained from the original mixtures, through a pre-whitening matrix $\mathbf{W}_{pre}$:

$$\mathbf{x} = \mathbf{W}_{pre}\mathbf{y}. \tag{8}$$

   The usual procedure to obtain the pre-whitening is to apply square-root factorization to the sample correlation matrix $\hat{\mathbf{R}}_y = \frac{1}{L}\sum_{k=1}^{L} \mathbf{y}_k \mathbf{y}_k^T$ (of mixtures)

$$\hat{\mathbf{R}}_y = \hat{\mathbf{R}}_y^{1/2} \hat{\mathbf{R}}_y^{T/2}, \tag{9}$$

   where $\hat{\mathbf{R}}_y^{1/2}$ is any rectangular square-root of $\hat{\mathbf{R}}_y$. Then the pre-whitening matrix is set as

$$\mathbf{W}_{pre} = (\hat{\mathbf{R}}_y^{1/2})^{\dagger}. \tag{10}$$

2. *Orthogonal Separation:* The output of the previous step is an uncorrelated vector. However, it is not necessarily an independent vector. An adaptive orthogonal separator $\mathbf{\Theta}$ is trained to separate sources from the uncorrelated mixtures. The overall separator output can be written as

$$\mathbf{z} = \mathbf{\Theta}\mathbf{x}. \tag{11}$$

   In the ICA approach, obtaining orthogonal separator is typically posed as an optimization problem

$$\begin{array}{ll} \text{maximize} & J(\mathbf{\Theta}) \\ \text{subject to} & \mathbf{\Theta}\mathbf{\Theta}^T = \mathbf{I} \end{array} \tag{12}$$

   reflecting the goal to achieve independence in $\mathbf{z}$. The objective function selected as a measure of independence or non-Gaussianity of the separator output $\mathbf{z}$, such as negentropy and (negative) mutual information [1]. A common objective function is the cummulative Kurtosis function which can be written as

$$J(\mathbf{\Theta}) = \sum_{k=1}^{p} K(z_k) \quad (13)$$

where the kurtosis operator is defined as

$$K(u) = E(u^4) - 3E(u^2)^2. \quad (14)$$

It is common to apply a two step iterative algorithm to obtain the solution of the problem in (12):

$$\underline{\mathbf{\Theta}}^{(k+1)} = \mathbf{\Theta}^{(k)} + \mu^{(k)} \nabla_{\mathbf{\Theta}} J(\mathbf{\Theta}^{(k)}) \quad (15)$$
$$\mathbf{\Theta}^{(k+1)} = M_O(\underline{\mathbf{\Theta}}^{(k)}) \quad (16)$$

where

- $\mu^{(k)}$ is the step size at the k^{th} iterations,
- M_O is the mapping to orthogonal set.

In the first step of the above iterative method, the usual gradient ascent update is applied. However, this operation would potentially move the $\mathbf{\Theta}$ out of the constraint set. The second step maps the resulting orthogonal separator to a "nearby" orthogonal matrix. Among the different algorithms for M_O, the most common ones are

- *QR decomposition based* M_O: QR decomposition is applied to the argument of M_O. The output is selected as the orthogonal Q component of the decomposition.
- *Polar decomposition based* M_O: Polar decomposition is applied to the argument of M_O, i.e., it is written as the product of a orthogonal matrix with a positive definite matrix. Then the orthogonal component is assigned the output of the mapping. This M_O has the nice feature that the resulting output is the closest orthogonal matrix (in Frobenius norm or induced 2-norm sense) to the argument of the mapping.

B. Parallel BCA Algorithms

As discussed in the introduction section, Bounded Component Analysis do not assume the sources to be independent but bounded with domain separability condition. Therefore, in the following treatment we'll allow sources to be dependent, however, to draw direct parallels between the convergence analysis of ICA and BCA, we do assume that the sources are uncorrelated. The uncorrelated source assumption is not required in BCA framework. In fact, it is possible to develop BCA algorithms that can successfully separate dependent sources [27].

In reference [28], the following optimization scheme is proposed to separate independent sources with finite support:

$$\begin{array}{ll} \text{minimize} & L(\mathbf{\Theta}, \mathbf{b}) \\ \text{subject to} & \mathbf{\Theta}\mathbf{\Theta}^T = \mathbf{I} \end{array} \quad (17)$$

where

$$L(\mathbf{\Theta}, \mathbf{b}) = \sum_{k=1}^{p} \|\mathbf{\Theta}_{k,:}\mathbf{x} + \mathbf{b}\|_\infty \quad (18)$$

and $\mathbf{x}$ is the vector of whitened mixtures. The same setting can be used as a BCA approach for uncorrelated (but not necessarily independent) sources. The corresponding iterative parallel algorithm can be written as

$$\mathbf{b}^{(i+1)} = \mathbf{b}^{(i)} - \mu_{\mathbf{b}}^{(i)} \sum_{k=1}^{p} \sum_{l \in I_k^{(i)}} \lambda_l^{(k,i)} sign(z_k^{(i)}(l)) \mathbf{e}_k \quad (19)$$

$$\underline{\mathbf{\Theta}}^{(i+1)} = \mathbf{\Theta}^{(i)} - \mu_{\mathbf{\Theta}}^{(i)} \sum_{k=1}^{p} \sum_{l \in \mathcal{I}_k^{(i)}} \lambda_l^{(k,i)} sign(z_k^{(i)}(l)) \mathbf{e}_k {\mathbf{x}_l}^T \quad (20)$$

$$\mathbf{\Theta}^{(i+1)} = M_O\{\underline{\mathbf{\Theta}}^{(i+1)}\}, \quad (21)$$

where

- $\mathbf{e}_k$ is the standard basis vector with all zeros except k^{th} entry with unity value,
- $\mathcal{I}^{(k,i)}$ corresponds to the set containing indexes of the peak magnitues for the output k and at iteration i,
- $\lambda_l^{(k,i)}$ are the convex combination coefficients with the property $\lambda_l^{(k,i)} \geq 0$ and $\sum_{l \in \mathcal{I}_k^{(i)}} \lambda_l^{(k,i)} = 1$,
- μ_b and μ_Θ are step sizes for the corresponding search variables.

IV. Convergence Analysis Results

The main difficulty regarding the convergence analysis of both parallel ICA and BCA algorithms is the non-convex structure of the constraint set, i.e., the set of orthogonal matrices, in the corresponding optimization settings in (12) and (17). In fact, the main obstacle for the tractability of the analysis is due to the existence of M_O, mapping to the nearby orthogonal matrix, at the last phase of both algorithms. However, we can still obtain useful results for the convergence especially for the polar factorization based M_O (which will be referred as M_O^P). Two important features of the polar factorization based M_O, which are the significant tools for the analysis are outlined by the following two lemmas [29]:

Lemma 1: For a full rank square matrix $\mathbf{A}$ and a orthogonal matrix $\mathbf{\Theta}$, $M_O^P(\mathbf{\Theta}\mathbf{A}) = \mathbf{\Theta} M_O^P(\mathbf{A})$.

Lemma 2: For a full rank square matrix $\mathbf{A}$, $M_O^P(\mathbf{A}) = \mathbf{I}$, if and only if, $\mathbf{A}$ is a positive-definite matrix.

The convergence analysis results for parallel ICA and BCA algorithms employing M_O^P are discussed in the following two subsections.

A. Convergence Results for Parallel ICA Algorithms

The fixed or stationary points of the algorithms are the potential convergence points, and hence, their detailed investigation play a critical role. The following theorem [29], provides a general characterization of these points for the parallel ICA algorithm in (15-16):

Theorem 1: For the algorithm in (15-16) employing M_O^P, and using sufficiently small step sizes to ensure full rank condition for $\underline{\mathbf{\Theta}}$, the overall mapping $\mathbf{G}_*$ is a stationary point if and only if there exists a symmetric matrix $\mathbf{S}$ for which $\nabla_{\mathbf{G}} J(\mathbf{G}_*) = \mathbf{G}_*\mathbf{S}$.

According to the above theorem, for a orthogonal matrix to qualify as a stationary point of the parallel ICA algorithm, the gradient of the objective function at that point should lie in the normal space of the orthogonal matrix set at that point. The above characterization can be applied to some detailed characterization of the Kurtosis based objective function in (13). The gradient of this objective function (with respect to overall mapping $\mathbf{G}$) is given by

$$\nabla_{\mathbf{G}} J(\mathbf{G}) = 4\mathbf{G}^{\odot 3}\mathbf{K} \tag{22}$$

where $\mathbf{K} = diag(\kappa_1, \kappa_2, \ldots, \kappa_p)$ is a diagonal matrix containing source kurtosis values, which area assumed to be positive, and $\mathbf{G}^{\odot k}$ stands for k^{th} Hadamard power of $\mathbf{G}$. Based on this gradient expression and the characterization in Theorem 1, we can write examples of the stationary points as

- *Perfect Separators*: For perfect separators in (7), $\nabla_{\mathbf{G}} J(\mathbf{G}) = 4\mathbf{D}^3\mathbf{P} = \mathbf{G}4\mathbf{D}^2\mathbf{PK}$. Since $4\mathbf{D}^2\mathbf{PK}$ is symmetric (in fact, diagonal), perfect separators are stationary points
- *Matrices with constant magnitude nonzero entries at each column*: For a matrix whose non-zero entries on on column j has fixed magnitude (β_j), if we define $\mathbf{B} = diag(\beta_1, \beta_2, \ldots, \beta_p)$, we can write $\nabla_{\mathbf{G}} J(\mathbf{G}) = 4\mathbf{G}\mathbf{B}^2\mathbf{K}$. Stationary point property is implied by the fact that $\mathbf{B}^2\mathbf{K}$ is a diagonal matrix.
- *Householder Matrices with a normal whose non-zero entries have constant magnitude:* Matrices in this class can be specified as

$$\mathbf{G} = \mathbf{I} - 2\mathbf{v}\mathbf{v}^T \tag{23}$$

where $\mathbf{v}$ is a vector with L non-zero entries and the non-zero entries have the magnitude $\frac{1}{\sqrt{L}}$. For $\mathbf{K} = \kappa\mathbf{I}$, the gradient can be written as

$$\begin{aligned} \nabla_{\mathbf{G}} J(\mathbf{G}) &= 4\kappa(\mathbf{I} - 2\mathbf{v}\mathbf{v}^T)^{\odot 3} \\ &= 4\kappa(\mathbf{I} + \frac{6(2-L)}{L} diag(\mathbf{v}^{\odot 2}) - \frac{8}{L^2}\mathbf{v}\mathbf{v}^T \end{aligned}$$

Therefore, we can write

$$\begin{aligned} &\mathbf{G}^T \nabla_{\mathbf{G}} J(\mathbf{G}) \\ &= 4\kappa(\mathbf{I} - 2\mathbf{v}\mathbf{v}^T)(\mathbf{I} + \frac{6(2-L)}{L} diag(\mathbf{v}^{\odot 2}) - \frac{8}{L^2}\mathbf{v}\mathbf{v}^T \\ &= 4\kappa(\mathbf{I} + \frac{6(2-L)}{L} diag(\mathbf{v}^{\odot 2}) - \frac{2L^2 - 12L + 16}{L^2}\mathbf{v}\mathbf{v}^T \end{aligned}$$

which is a symmetric matrix, and therefore, confirms stationary point property of the special Householder matrices. The classification of the special Householder matrices is given by the following theorem [29]:

Theorem 2: The special Householder based stationary points are classified as follows:

  - $L = 1$ and $L = 2$ corresponds to Global Maxima.
  - $L = p = 4$ corresponds to Global Minima.
  - All remaining cases are Saddle Points.

Another useful convergence result concerns the general objective functions $J(\mathbf{\Theta})$, which is stated in the following theorem:

Theorem 3: The algorithm in (15-16), for sufficiently small step sizes (μ's), monotonically (in objective function value) to a stationary point of the algorithm.

This result is deduced from the Theorem 4 in [29], which proves the monotonic convergence for the fixed point algorithm and the convex objective functions, and its extension for non-convex objective functions. The bound for the step size to satisfy the sufficiency condition depends on the minimum eigenvalue of the Hessian matrix at the iteration point.

B. Convergence Results for Parallel BCA Algorithms

The dual of the stationary point characterization in Theorem 1 for parallel BCA algorithms in (19-21) is given by the following theorem [28]:

Theorem 4: For the algorithm in (19-21) employing M_O^P, and using sufficiently small step sizes to ensure full rank condition for $\underline{\mathbf{\Theta}}$, the overall mapping $\mathbf{G}_*$ is a stationary point if and only if $\mathbf{G}_*^T sign(\mathbf{G}_*)$ is a symmetric matrix.

Based on this theorem, we can show that the stationary points listed for the parallel ICA algorithm with Kurtosis based objective function are also stationary points of the parallel BCA algorithm in (19-21):

- *Perfect Separators*: For perfect separators in (7), $sign(\mathbf{G}) = \mathbf{G}$, therefore, $\mathbf{G}^T sign(\mathbf{G}) = \mathbf{I}$ is clearly symmetric.
- *Matrices with constant magnitude nonzero entries at each column*: In this case $sign(\mathbf{G}) = \mathbf{G}\mathbf{B}^{-1}$. Therefore, $\mathbf{G}^T sign(\mathbf{G}) = \mathbf{B}^{-1}$, which is diagonal (hence symmetric).
- *Householder Matrices with a normal whose non-zero entries have constant magnitude:* We can show that

$$\mathbf{G}^T sign(\mathbf{G}) = \mathbf{I} + L diag(\mathbf{v}^{\odot 2}) - \frac{L-4}{L}\mathbf{v}\mathbf{v}^T \tag{24}$$

which is a symmetric matrix. The classification in Theorem 2 can be shown to be valid for the parallel BCA algorithm [30].

As a result, we can see that the two step parallel ICA and BCA algorithms share the same set of critical points. However, we should warn that there is no result (yet) that this shared set of stationary points is the comprehensive set of stationary points for any of the two algorithms.

V. Conclusion

In this article, we comparatively reviewed some fundamental convergence analysis results corresponding to two different parallel BSS schemes. Among them, Independent Component Analysis is the well established and most popular scheme whereas the other approach, namely the Bounded Component Analysis, is a recently introduced and a promising technique.

Main attractive property of the BCA scheme is its ability to separate dependent signals as well as independent ones. The main message of the article concerns the parallelism between the convergence structure of these alternative BSS methods. In fact, they share the same (identified) set of stationary points.

There are several open questions to be answered such as whether the presented set of stationary points is comprehensive for both schemes or whether there are additional stationary points which are not identified yet. Furthermore, despite the duality between the two schemes in terms of stationary points, there is no monotonic convergence result for the presented BCA approach which would be dual of the ICA result in Theorem 3.

REFERENCES

[1] P. Comon and C. Jutten, *Handbook of Blind Source Separation: Independent Component Analysis and Applications*. Academic Press, 2010.

[2] P. Comon, "Independent component analysis, a new concept?"

[3] J. F. Cardoso, "Higher-order contrasts for independent component analysis: Algorithms and applications," *Neural Computation*, vol. 11, no. 1, pp. 157–192, March 1999.

[4] A. Hyvarinen and E. Oja, "Independent component analysis: Algorithms and applications," *Neural Networks*, vol. 13, no. 4-5, pp. 411–430, 2000.

[5] D.-T. Pham, "Mutual information approach to blind separation of stationary sources," *IEEE Trans. on Signal Processing*, vol. 48, pp. 1935 – 1946, July 2002.

[6] D.-T. Pham and P. Garat, "Blind separation of mixture of independent sources through a quasi-maximum likelihood approach," *IEEE Trans. on Signal Processing*, vol. 45, pp. 1712 – 1725, July 1997.

[7] A. Hyvärinen, "Fast and robust fixed-point algorithms for independent component analysis," *IEEE Trans. on Neural Networks*, vol. 10, no. 3, pp. 626–634, 1999.

[8] A. Belouchrani, K. Abed-Meraim, and J. F. Cardoso, "A blind source separation technique using second order statistics," *IEEE Transactions on Signal Processing*, vol. 45, no. 2, pp. 434–444, February 1997.

[9] 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 1997.

[10] W. A. Gardner, A. Napolitano, and L. Paura, "Cyclostationarity: Half a century of research," *Signal Processing*, vol. 86, no. 4, pp. 639–697, April 2006.

[11] P. Georgiev, F. Theis, and A. Cichocki, "Sparse component analysis and blind source separation of underdetermined pictures," *IEEE Transactions on Neural Networks*, vol. 16, no. 4, pp. 992–996, July.

[12] O. Yilmaz and S. Rickard, "Blind separation of speech mixtures via time-frequency masking," *IEEE Transactions on Signal Processing*, vol. 52, no. 7, pp. 1830–1847, July 2004.

[13] A. 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.

[14] P. Callon, A. Chevreuil, 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.

[15] 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.

[16] D.-T. Pham, "Blind separation of instantenaous mixtrures of sources based on order statistics," *IEEE Trans. on SP*, vol. 48, no. 2, pp. 363–375, February 2000.

[17] 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.

[18] 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.

[19] 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.

[20] C. 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.

[21] T. M. Oktem, A. T. Erdogan, and A. Demir, "Adaptive receiver structures for fiber communication systems employing polarization division multiplexing," *IEEE/OSA Journal of Lightwave Technology*, vol. 27, no. 3, pp. 5394–5404, December 2009.

[22] ——, "Adaptive receiver structures for fiber communication systems employing polarization division multiplexing: High symbol rate case," *IEEE/OSA Journal of Lightwave Technology*, vol. 28, no. 10, pp. 1536–1546, May 2010.

[23] 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.

[24] 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.

[25] F. Vrins, J. A. Lee, and M. Verleysen, "A minimum-range approach to blind extraction of bounded sources," *IEEE Transactions on Neural Networks*, vol. 18, no. 3, pp. 809–822, May 2007.

[26] S. Cruces, "Bounded component analysis of linear mixtures: A criterion for minimum convex perimeter," *IEEE Trans. on SP.*, vol. 58, no. 4, pp. 2141–2154, 2010.

[27] A. T. Erdogan, "A family of bounded component analysis algorithms," *submitted to ICASSP 2012 Conference*.

[28] ——, "Adaptive algorithm for the blind separation of sources with finite support," *Proceedings of EUSIPCO 2008, Lausanne*, August 2008.

[29] ——, "On the convergence of ICA algorithms using symmetric orthogonalization," *IEEE Trans. on Signal Processing*, vol. 57, no. 06, pp. 2209–2221, June 2009.

[30] ——, "Bounded component analysis algorithms for the separation of uncorrelated sources," *Under preperation*.