A New Proof of a Contrast Function for Bounded Component Analysis and Further Analysis

Wei Gao[a,*], Shen Fan[b], Roberto Togneri[a], Victor Sreeram[a]

[a]*School of Electrical Electronic and Computer Engineering, The University of Western Australia, Perth 6009, Australia.*
[b]*College of Science, China University of petroleum (Beijing), Beijing 102200, China*

Abstract

Bounded Component Analysis (BCA) solves the Blind Source Separation (BSS) problem based on geometric assumptions. This paper introduces a new proof of a BCA contrast function, derived from elementary matrices, Gauss-Jordan elimination and convex geometry. The new proof and further analysis provide additional insight into a key assumption of BCA. In addition, an interpretation is presented to clarify one of the limitations of the instantaneous BCA algorithm. Experiments on audio sources support our analysis.

Keywords: bounded component analysis, blind source separation, independent component analysis, elementary matrix, convex geometry, audio signals.

1. Introduction

As the name "blind" suggests, Blind source separation (BSS) aims to recover the sources from mixtures of the sources only, without prior information of the sources and the way the sources were mixed [1]. The "blind" feature not only leads to a broad variety of applications in practice, such as

*Corresponding author

Email addresses: wei.gao@research.uwa.edu.au (Wei Gao), fans@cup.edu.cn (Shen Fan), roberto.togneri@uwa.edu.au (Roberto Togneri), victor.sreeram@uwa.edu.au (Victor Sreeram)

[1]The work of Shen Fan was supported by Science Foundation of China University of Petroleum, Beijing (No.YJRC-2011-14, No.FZG-2011-01).

speech processing, digital communications, image processing and chemometrics, etc., but also makes BSS an exciting challenge as an ill-posed problem [2]. Prior information is generally replaced by some assumptions. The more stringent the assumptions, the narrower the applicability [3], but the easier the BSS problem. There is no BSS algorithm that can be universally used to solve all problems.

A wide variety of assumptions and corresponding algorithms have been developed in different application fields. Independent Component Analysis (ICA) algorithms can be considered as the representation of BSS solutions in the 1990's, proposed in the field of neural networks [4] [5]. Rooted in statistically independent or uncorrelated assumptions of the sources, numerous ICA algorithms of higher order statistics (HOS) and second order statistics (SOS) were developed [6]. For example, SOBI [7], STOTD [8] and JADE [9][10] combine Jacobi rotation with second-order, third-order and fourth-order cumulants, respectively. For computational and conceptual simplicity, BSS solutions were obtained by optimizing a contrast function, for example FastICA [11], a classical and implementable ICA algorithm.

ICA algorithms were further developed by incorporating geometric techniques into the implementation of the BSS algorithm. For nonstatioanary sources, a new joint diagonalization procedure was established based on the principles of maximum likelihood and minimum mutual information [12]. To deal with data contamination, the gamma-ICA method with better robustness was proposed, and implemented by a geometrical framework based on gradient flows on a special orthogonal matrix [13]. The minimum-range approach investigates bounded sources, and geometric interpretation is exploited for contrast maximization over the group of special orthogonal matrices [14]. Similarly, under a certain boundedness assumption of sources, minimization of the infinity norm approach is centered around the basic geometric fact of independent vectors [15]. The restrictions of statistical independence of the sources can be relaxed to partial correlation, a constrast function was proposed by combining geometric techuniques [16].

However, some realistic sources are dependent or correlated. To replace the restrictions of statistical independence, various assumptions and algorithms were proposed, such as Non-negative Matrix Factorization (NMF) and Sparse Component Analysis (SCA) [2].

Compared with NMF and SCA, Bounded Component Analysis (BCA) is a relatively new avenue for BSS [2] without the assumption of independence of the sources. The BCA proposed by Cruces [17], relies on the assump-

tions of compactness and Cartesian decomposition of the convex support of the vector of the sources. Still based on the convex hull of the sources, [18] developed BCA algorithms by considering the underdetermined mixtures. An alternative framework proposed by Erdogan utilises the principal hyper-ellipsoid and the bounding hyper-rectangle of the sources [19], which supports Cruces's idea that BCA can be considered as a more general approach, covering ICA as a special case for bounded sources [17]. Then, Erdogan's BCA algorithms were further developed for separating convolutive mixtures [20][21]. Furthermore, sparse BCA algorithm was proposed to consider the sparsity of the sources [22]. Recently, a stationary point for instantaneous BCA algorithms was analysed [23]. Although BCA algorithms have attracted much attention and developed quickly, there is still a question whether BCA should be considered as a more general approach than ICA. In this paper, we will show that BCA should be considered as a complementary approach to ICA, rather than a more general approach covering ICA.

Cruces [17] assumed the following three properties to ensure the separability:

P1) *Compactness and nondegeneracy of the sources*: all the sources are nondegenerate[2] random variables of compact support.

P2) *Cartesian decomposition of the convex support of the sources*: $\mathbb{S}_{\check{S}} = \mathbb{S}_{\check{S}_{1,:}} \otimes \cdots \otimes \mathbb{S}_{\check{S}_{p,:}}$, where $\otimes$ denotes Cartesian product, $\mathbb{S}_{\check{S}}$ denote the convex hull of the support of the sources, and $\mathbb{S}_{\check{S}_{i,:}}$ denote the convex hull of the support of the ith source.

P3) *Lossless mixing*: the mixing matrix is full-column rank.

On the basis of **P1)-P3)**, in [19] one more assumption **A1)** was introduced: S contains the vertices of its bounding hyper-rectangle.

Although [19] have provided a proof of the contrast function (4), it lends little insight into the condition when assumption **A1)** is not satisfied. This paper not only presents a new proof of the contrast function (4) in Section 2, but also in Section 3 provides a more in-depth analysis and insight into the following five aspects regarding BCA:

(i) Theoretical discussion on the link between **A1)** and **P2)**;

[2]A random variable can be considered degenerate if the support of its p.d.f consist in a single point.

(ii) Although BCA is named by the boundedness of the sources **P1)**, the key and stringent assumption of BCA is Cartesian decomposition of the sources, i.e. **P2)**;
(iii) **A1)** is likely to be stringent in some practical applications;
(iv) Interpretation on why the BCA algorithm suffers when **A1)** is not satisfied;
(v) Experiments show how the instantaneous BCA algorithm suffers when applied to blind audio source separation.

The analysis and experiments provide a comprehensive recognition of BCA, not only advantages but also limitations. Equipped with the comprehensive recognition of BCA, readers have more opportunities to select a suitable BSS algorithm.

The remainder of this paper is organised as follows: Section 2 presents the BCA models and notation, as well as some preliminaries; Section 3 provides our proof of Erdogan's BCA contrast function; the BCA assumptions are further analysed in Section 4, and supported by numerical experiments; finally, our conclusion is presented in Section 5.

2. Background

2.1. Models and Notation

Consider an instantaneous real-valued BSS model without noise. Observed signals $Y \in \mathbb{R}^{q\times L}$ are linear instantaneous mixtures of the sources $S \in \mathbb{R}^{p\times L}$ such that $Y = HS$, where $H \in \mathbb{R}^{q\times p}$ denotes the mixing matrix. Here L, q and p indicate the number of samples, mixtures and sources, respectively. Note that from the theoretical point of view L can be assumed to be infinite, but in practical applications L is a finite integer. BSS is concerned with finding a separating matrix $W \in \mathbb{R}^{p\times q}$ to obtain the recovered sources $Z = WY = GS$, where $G = WH$ represents the overall mapping. The L samples of the ith source are expressed as $S_{i,:}$, that is the ith row of S. Similarly, $Z_{i,:}$ and $(\cdot)_{i,:}$ denote the ith row of Z and the enclosed matrix, respectively. Each sample of the sources are expressed as $S_{:,j}$, that is the jth column of S. Similarly, $(\cdot)_{:,j}$ denote the jth column of the enclosed matrix.

Let

$$\hat{\mathbf{I}}(S) = \begin{bmatrix} \min_{k\in\{1,\cdots,L\}} S_{1,k} \\ \vdots \\ \min_{k\in\{1,\cdots,L\}} S_{p,k} \end{bmatrix} \tag{1}$$

$$\hat{\mathbf{u}}(S) = \begin{bmatrix} \max_{k\in\{1,\cdots,L\}} S_{1,k} \\ \vdots \\ \max_{k\in\{1,\cdots,L\}} S_{p,k} \end{bmatrix} \tag{2}$$

denote the vectors containing minimum and maximum values for the rows of S, respectively. The bounding hyper-rectangle of S can be defined as

$$\widehat{B}(S) = \left\{ \mathbf{q} : \hat{\mathbf{l}}(S) \leqslant \mathbf{q} \leqslant \hat{\mathbf{u}}(S) \right\} \tag{3}$$

Thus, $\widehat{B}(S)$ is defined as the minimum volume box covering all samples of S and aligning with the coordinate axes. Vertices of $\widehat{B}(S)$ are determined by the minimum and maximum of each row of S. Similarly, $\widehat{B}(Z)$ denotes the bounding hyper-rectangle of Z.

Additionally, $\mathrm{cov}(\cdot)$, $\mathrm{abs}(\cdot)$ and $\det(\cdot)$ express the covariance matrix, the absolute value and the determinant, respectively. The center of the principal hyper-ellipsoid of S is given by the sample mean of S, which is defined as

$$\hat{\mu}(S) = \frac{1}{L} \sum_{k=1}^{L} S_{:,k}.$$

The principal semiaxes directions are determined by the eigenvectors of the sample covariance matrix of S as

$$\mathrm{cov}(S) = \frac{1}{L} \sum_{k=1}^{L} (S_{:,k} - \hat{\mu}(S))(S_{:,k} - \hat{\mu}(S))^T.$$

Then, principal hyper-ellipsoid of S can be defined as

$$\widehat{\varepsilon}(S) = \{ \mathbf{q} : (\mathbf{q} - \hat{\mu}(S))^T \mathrm{cov}(S)^{-1} (\mathbf{q} - \hat{\mu}(S)) \leqslant 1 \}.$$

Under the above four assumptions, Erdogan [19] proposed a family of contrast functions based on:

$$\text{maximize} \left\{ \frac{\mathrm{vol}[\widehat{\varepsilon}(Z)]}{\mathrm{vol}[\widehat{B}(Z)]} = \frac{\mathrm{vol}[\widehat{\varepsilon}(GS)]}{\mathrm{vol}[\widehat{B}(GS)]} \right\}, \tag{4}$$

where $\mathrm{vol}[\cdot]$ denotes the volume. The volume of the principal hyper-ellipsoid of Z is:

$$\mathrm{vol}[\widehat{\varepsilon}(Z)] = C_p \sqrt{\det(\mathrm{cov}(Z))} = \mathrm{vol}[\widehat{\varepsilon}(GS)] = \mathrm{abs}(\det(G)) \mathrm{vol}[\widehat{\varepsilon}(S)], \tag{5}$$

where $C_p = \pi^{\frac{p}{2}}/\Gamma\left(\frac{p}{2}\right)$, and $\Gamma(\cdot)$ denotes the Gamma function. The volume of the bounding hyper-rectangle of Z can be expressed compactly as:

$$\mathrm{vol}[\widehat{B}(Z)] = \prod_{i=1}^{p} \{\max[Z_{i,:}] - \min[Z_{i,:}]\} = \mathrm{vol}[\widehat{B}(GS)]. \tag{6}$$

2.2. Preliminaries

In linear algebra, there are three types of elementary matrices E_1, E_2 and E_3, which are defined from the identity matrix I as follows:

$$E_1(I_u \leftrightharpoons I_v) = \begin{bmatrix} 1 & & & & & & \\ & \cdot & & & & & \\ & & 0 & & 1 & & \\ & & & \cdot & & & \\ & & 1 & & 0 & & \\ & & & & & \cdot & \\ & & & & & & 1 \end{bmatrix} \tag{7}$$

where I_u (the uth row of I) is switched with I_v (the vth row of I);

$$E_2(hI_u) = \begin{bmatrix} \cdot & & & & \\ & 1 & & & \\ & & h & & \\ & & & 1 & \\ & & & & \cdot \end{bmatrix} \tag{8}$$

where I_u is multiplied by a non-zero scalar $h \neq 0$;

$$E_3(kI_v \to I_u) = \begin{bmatrix} \cdot & & & \\ & 1 & & k \\ & & \cdot & \\ & & & 1 \\ & & & & \cdot \end{bmatrix} \tag{9}$$

where $k \neq 0$ and I_v multiplied by k is added to I_u. For brevity of notation, hereinafter elementary matrices are of appropriate dimension for matrix multiplications, and the parameters for the elementary matrices are omitted. Left multiplication by E_1, E_2 and E_3 correspond to reflection mapping, scale mapping and shear mapping, respectively.

A convex hull of a set $X = \{X_{:,1}, \cdots, X_{:,L}\}$, where $X_{:,i} \in \mathbb{R}^N$ denotes a point, can be defined as

$$Conv(X) = \left\{ \sum_{i=1}^{L} X_{:,i}\theta_i \Big| \sum_{i=1}^{L} \theta_i = 1 \wedge (\forall i : \ \theta_i \geqslant 0) \right\}$$

Roughly speaking, $Conv(X)$ is the smallest convex set that contains X. For a compact finite set X, if A is a linear mapping (equivalent to a linear matrix in linear algebra), then $Conv(A(X)) = A(Conv(X))$ and $\text{vol}[Conv(A(X))] = \text{abs}(\det(A))\text{vol}[Conv(X)]$ [24][25].

3. New Proof of Erdogan's BCA

The new proof presented here assumes real-valued signals. The extension to complex-valued signals is relatively straightforward, albeit cumbersome, requiring the $\max[\cdot]$ and $\min[\cdot]$ operations to be carried out on the real-part and imaginary-part of signals separately. This essentially treats complex-valued signals as paired real-valued signals. In this section, we first analyse the mixing matrix and the overall mapping. Then our proof is presented in three steps. Finally, our remarks following our proof are presented.

3.1. Analysis of the Overall Mapping

Note that the geometric concepts $\widehat{B}(S)$ and $Conv(S)$ are formulated by viewing the matrix S as a set composed of the columns of S, i.e. $S = \{S_{:,1}, \cdots, S_{:,j}, \cdots, S_{:,L}\}$, where $S_{:,j}$ represents a column of the matrix S. This view has been the basis of a scatter plot in the BSS problem [2]. We exploit this view and geometric concepts to analyse the BCA contrast function.

Under **P1)**, "compactness" implies that each source $S_{i,:}$ attains a finite maximum and minimum; "nondegeneracy" implies that for each source $S_{i,:}$, its maximum is greater than its minimum. It follows that $\max[S_{i,:}] - \min[S_{i,:}] > 0, i \in \{1, \cdots, p\}$. Hence, substituting E_1 in (7) into (6),i.e., $G = E_1$, we get for any E_1

$$\begin{aligned}
\text{vol}\left[\widehat{B}(E_1S)\right] &= \prod_{i=1}^{p} \{\max[(E_1S)_{i,:}] - \min[(E_1S)_{i,:}]\} \\
&= \prod \{\max[S_{i,:}] - \min[S_{i,:}]\}, i = 1, \cdots, v, \cdots, u, \cdots, p \\
&= \prod \{\max[S_{i,:}] - \min[S_{i,:}]\}, i = 1, \cdots, u, \cdots, v, \cdots, p \\
&= \text{vol}[\widehat{B}(S)].
\end{aligned}$$

Given $\text{abs}(\det(E_1)) = 1$, we can state more generally that

$$\text{vol}\left[\widehat{B}(E_1 S)\right] = \text{abs}(\det(E_1))\ \text{vol}[\widehat{B}(S)],\ \forall E_1. \tag{10}$$

In the same way, substituting E_2 in (8) into (6), and noting $\text{abs}(\det(E_2)) = \text{abs}(h)$, we get

$$\text{vol}\left[\widehat{B}(E_2 S)\right] = \text{abs}(h)\ \text{vol}[\widehat{B}(S)] = \text{abs}(\det(E_2))\ \text{vol}[\widehat{B}(S)],\ \forall E_2. \tag{11}$$

Note that the elementary matrix has the property of full rank. Thus, if we consider an elementary matrix as the simplest mixing matrix H or the simplest overall mapping G, the equalities of (10) and (11) rely on **P1)** and **P3)**.

Under **A1)** and **P1)**, the bounding hyper-rectangle $\widehat{B}(S)$ is a nondegenerate p-dimensional hyper-rectangle with edges aligning with coordinate axes in geometry. Under **P3)**, G is a full-rank square matrix, and it follows that G is invertible. Therefore, under G the bounding hyper-rectangle $\widehat{B}(S)$ is mapped to a nondegenerate p-dimensional hyper-parallelogram $G\widehat{B}(S)$, whose edges may not align with coordinate axes. We consider the constraints on G that make "align with coordinate axes" preserved, in the following Lemma.

Lemma 1. *Under an invertible linear G, if $G\widehat{B}(S)$ is also a bounding hyper-rectangle with edges aligning with coordinate axes, then there exist D and P such that $G = DP$, where D is an invertible square diagonal matrix, and P is an invertible square permutation matrix.*

Proof. The p column vectors $e_1 = [1, 0, \cdots, 0]^T, \cdots, e_p = [0, \cdots, 0, 1]^T$ denote p unit-coordinate vectors of $\mathbb{R}^p$, such that e_k is the kth column of an identity matrix $I \in \mathbb{R}^{p \times p}$. Since the edges of the bounding hyper-rectangle are aligned with the coordinate axes, for any vertex V_0 of $\widehat{B}(S)$, there exists an edge $\overrightarrow{V_0 V_k}$ starting from V_0 to another vertex V_k that is parallel to e_k, $k = 1, \cdots, p$. Let a_k denote the length of $\overrightarrow{V_0 V_k}$, and we then have

$$\overrightarrow{V_0 V_k} = a_k e_k,\ k = 1, \cdots, p, \tag{12}$$

where $a_k \neq 0$ by **P1)**.

Denote the images of $V_0, V_1, \cdots, V_p$ under G as $V_0^{'}, V_1^{'}, \cdots, V_p^{'}$, which are still the vertices of $G\widehat{B}(S)$, given G is an invertible linear mapping by **P3)**. Then, the mapped edge

$$\overrightarrow{V_0^{'} V_k^{'}} = G\overrightarrow{V_0 V_k},\ k = 1, \cdots, p. \tag{13}$$

If $G\widehat{B}(S)$ is also a p-dimensional bounding hyper-rectangle, which means each edge of $G\widehat{B}(S)$ is aligned with a coordinate axis, there exists e_{i_k} parallel to $\overrightarrow{V_0^{'}V_k^{'}}$, where $i_k \in \{1, \cdots, p\}$ is the index of the unit-coordinate vector, i.e., the linear mapping G maps e_k to e_{i_k}. Denote b_{i_k} the length of the edge $\overrightarrow{V_0^{'}V_k^{'}}$, so

$$\overrightarrow{V_0^{'}V_k^{'}} = b_{i_k} e_{i_k},\ k = 1, \cdots, p. \tag{14}$$

Since G is an invertible matrix, $b_{i_k} \neq 0, \forall k \in \{1, \cdots, p\}$, and $i_1, i_2 \cdots i_p$ is a permutation of $1, 2, \cdots, p$.

Combining (12),(13) and (14) together, we have $G\overrightarrow{V_0V_k} = a_k G e_k = b_{i_k} e_{i_k},\ k = 1, \cdots, p$, and hence we get $Ge_k = \frac{b_{i_k}}{a_k} e_{i_k}$. Given e_k is the kth column of I, that is $[e_1, e_2, \cdots, e_p] = I$, we have

$$G = G[e_1, e_2, \cdots, e_p] = \begin{bmatrix} \frac{b_{i_1}}{a_1} & & & 0 \\ & \cdot & & \\ & & \frac{b_{i_k}}{a_k} & \\ & & & \cdot \\ 0 & & & \frac{b_{i_p}}{a_p} \end{bmatrix} [e_{i_1}, \cdots, e_{i_p}],$$

Let D denote the square diagonal matrix

$$\begin{bmatrix} \frac{b_{i_1}}{a_1} & & & 0 \\ & \cdot & & \\ & & \frac{b_{i_k}}{a_k} & \\ & & & \cdot \\ 0 & & & \frac{b_{i_p}}{a_p} \end{bmatrix},$$

and P denote the square permutation matrix $[e_{i_1}, \cdots, e_{i_p}]$. Since G is an invertible matrix, $b_{i_k} \neq 0, \forall k \in \{1, \cdots, p\}$, and $i_1, i_2 \cdots i_p$ is a permutation of $1, 2, \cdots, p$. Thus, $G = DP$ where D is an invertible square diagonal matrix, and P is an invertible square permutation matrix. □

3.2. New Proof of the BCA Contrast Function

Our proof of Erdogan's contrast function (4) can be divided into three steps: Step1 proves that if G is a perfect solution, then $\text{vol}[\widehat{B}(GS)] = \text{abs}(\det(G))\text{vol}[\widehat{B}(S)]$; Step2 proves that if the condition in Step1 holds, then G is a perfect solution; Step3 combines Step1 and Step2 together as proof of the contrast function (4).

Proof. **Step1**

Under **P3)**, the overall mapping G is a full-rank square matrix, which can be classified into two classes: the perfect solution denoted by $\bar{G}$, and the imperfect solution denoted by $\underline{G}$. Due to unavoidable ambiguities of BSS, if G is up to row permutation and scaling from an identity matrix, G is classified as a perfect solution. It follows that any perfect solution $\bar{G} = DP$ [2], where D is an invertible square diagonal matrix, and P is an invertible square permutation matrix.

According to Gauss-Jordan elimination, D can be simply factorized as a product of finite E_2 factors, $D \triangleq \prod_{f_2=1}^{F_2} E_2^{f_2}$, where the superscript f_2 denotes the f_2th factor, and F_2 is a non-negative integer denoting the total number of E_2 factors. Note that $\triangleq$ denotes simply factorized, which requires that every elementary matrix factor impacts different entries of I. In other words, any identity matrix and any elementary matrix should not be further factorized to two or more elementary matrices.

Similarly, P can be simply factorized as a product of finite E_1 factors, $P \triangleq \prod_{f_1=1}^{F_1} E_1^{f_1}$, where the superscript f_1 denotes the f_1th factor, and F_1 is a non-negative integer denoting the number of E_1 factors. We can state more generally that, if and only if $\bar{G}$ is a perfect solution,

$$\bar{G} \triangleq \prod_{f=0}^{F} E_i^f, i \in \{1, 2\}, \tag{15}$$

where the superscript f denotes the fth elementary matrix factor, and $F = F_1 + F_2$ denotes the total number of elementary matrix factors. If $F_1 = 0$ and $F_2 = 0$, $\bar{G} = I$.

By Gauss-Jordan elimination, the total number of simply factorized factors of G is finite. After finite transformations, the equality of (10) and (11) will still hold, under **P1)** and **P3)**. For the perfect solution $\bar{G}$, substituting (15) into (6), and simplifying sequentially by (10) and (11) together, we get

$$\begin{aligned} \mathrm{vol}\left[\widehat{B}(\bar{G}S)\right] &= \prod_{f_2=1}^{F_2} \mathrm{abs}(\det(E_2^{f_2})) \prod_{f_1=1}^{F_1} \mathrm{abs}(\det(E_1^{f_1}))\mathrm{vol}[\widehat{B}(S)] \\ &= \mathrm{abs}(\det(D))\ \mathrm{abs}(\det(P))\ \mathrm{vol}[\widehat{B}(S)] = \mathrm{abs}(\det(\bar{G}))\mathrm{vol}[\widehat{B}(S)]. \end{aligned} \tag{16}$$

Step2

Under **P1)**, by the definition in (1) (2) and (3), $\widehat{B}(S)$ is an intersection of a lower half-space $\{\mathbf{q} \leqslant \hat{\mathbf{u}}(S)\}$ and an upper half-space $\{\hat{\mathbf{l}}(S) \leqslant \mathbf{q}\}$. Thus, $\widehat{B}(S)$ is a compact convex set that contains S. Since the convex hull $Conv(S)$ is the intersection of all convex sets containing S, $Conv(S) \subseteq \widehat{B}(S)$. With the addition of **A1)**: S contains the vertices of its bounding hyper-rectangle, and these vertices are the maximum and minimum values of each row of S. It follows that the vertices of $\widehat{B}(S)$ are also the vertices of $Conv(S)$. Hence, the compact convex set $\widehat{B}(S)$ is the smallest convex set that contains S, and $\widehat{B}(S)$ is also the convex hull of S. This means that under **P1)** and **A1)**,

$$\widehat{B}(S) = Conv(S). \tag{17}$$

By definition, $\widehat{B}(GS)$ is a compact convex set that contains GS, so

$$Conv(GS) \subseteq \widehat{B}(GS). \tag{18}$$

Since G is a linear transformation, we have

$$GConv(S) = Conv(GS). \tag{19}$$

Combining (17) (18) and (19) together, we get

$$G\widehat{B}(S) = GConv(S) = Conv(GS) \subseteq \widehat{B}(GS). \tag{20}$$

Then, $\text{vol}[G\widehat{B}(S)] \leqslant \text{vol}[\widehat{B}(GS)]$, where the " = " is of interest. We use contradiction to prove that if $\text{vol}[G\widehat{B}(S)] = \text{vol}[\widehat{B}(GS)]$, then $G\widehat{B}(S) = \widehat{B}(GS)$.

Assume to the contrary that when $\text{vol}[G\widehat{B}(S)] = \text{vol}[\widehat{B}(GS)]$, $G\widehat{B}(S) \neq \widehat{B}(GS)$. By (20), the contrary assumption follows that when $\text{vol}[G\widehat{B}(S)] = \text{vol}[\widehat{B}(GS)]$, $G\widehat{B}(S) \subset \widehat{B}(GS)$. Given $G\widehat{B}(S) \subset \widehat{B}(GS)$, there must exist a vertex V of $\widehat{B}(GS)$ such that $V \notin G\widehat{B}(S)$. On one hand, there exists a closed hyperball $B(V, r)$ of radius $r > 0$ centred at V such that $B(V, r) \cap G\widehat{B}(S) = \emptyset$, which follows that $\text{vol}[B(V, r) \cap G\widehat{B}(S)] = 0$. On the other hand, V is a vertex of the hyper-rectangle $\widehat{B}(GS)$, so $\text{vol}[B(V, r) \cap \widehat{B}(GS)] > 0$. This leads to $\text{vol}[G\widehat{B}(S)] < \text{vol}[\widehat{B}(GS)]$, which is a contradiction of the above assumption that $\text{vol}[G\widehat{B}(S)] = \text{vol}[\widehat{B}(GS)]$. Thus, we can prove if $\text{vol}[G\widehat{B}(S)] = \text{vol}[\widehat{B}(GS)]$, then $G\widehat{B}(S) = \widehat{B}(GS)$.

Furthermore, if $G\widehat{B}(S) = \widehat{B}(GS)$, then $\text{vol}[G\widehat{B}(S)] = \text{vol}[\widehat{B}(GS)]$. Hence, we can conclude $\text{vol}[G\widehat{B}(S)] = \text{vol}[\widehat{B}(GS)]$, if and only if $G\widehat{B}(S) = \widehat{B}(GS)$.

In addition, since G is an invertible linear mapping and $\widehat{B}(S)$ is a convex set, we have $\text{vol}[G\widehat{B}(S)] = \text{abs}(\det(G))\text{vol}[\widehat{B}(S)]$. Hence, we have

$$\text{abs}(\det(G))\text{vol}[\widehat{B}(S)] = \text{vol}[G\widehat{B}(S)] \leqslant \text{vol}[\widehat{B}(GS)], \tag{21}$$

where the second " = " holds if and only if $G\widehat{B}(S) = \widehat{B}(GS)$. It follows that $\text{abs}(\det(G))\text{vol}[\widehat{B}(S)] = \text{vol}[\widehat{B}(GS)]$, if and only if $G(\widehat{B}(S)) = \widehat{B}(GS)$. Moreover, $G\widehat{B}(S) = \widehat{B}(GS)$ means $G\widehat{B}(S)$ is a bounding hyper-rectangle. By Lemma 1, for G such that $\text{abs}(\det(G))\text{vol}[\widehat{B}(S)] = \text{vol}[\widehat{B}(GS)]$, there exists $G = DP$, which is a perfect solution. Therefore, we can conclude that G such that $\text{vol}[\widehat{B}(GS)] = \text{abs}(\det(G))\text{vol}[\widehat{B}(S)]$ is a perfect solution.

Step3

In **Step1**, (16) indicates a perfect solution $\bar{G}$ is such that $\text{vol}[\widehat{B}(\bar{G}S)] = \text{abs}(\det(\bar{G}))\text{vol}[\widehat{B}(S)]$. Conversely, **Step2** proves that when $\text{vol}[\widehat{B}(GS)] = \text{abs}(\det(G))\text{vol}[\widehat{B}(S)]$, G is a perfect solution. Thus, we can conclude that if and only if G is a perfect solution that the second " = " of (21) will hold. Combining (21) and (5) into (4), we have for any G

$$\frac{\text{vol}[\widehat{\varepsilon}(GS)]}{\text{vol}[\widehat{B}(GS)]} \leqslant \frac{\text{abs}(\det(G))\text{vol}[\widehat{\varepsilon}(S)]}{\text{abs}(\det(G))\text{vol}[\widehat{B}(S)]} = \frac{\text{vol}[\widehat{\varepsilon}(S)]}{\text{vol}[\widehat{B}(S)]}, \tag{22}$$

where if and only if G is a perfect solution that the first " = " will hold. The right side of (22) is independent of G, and only dependent on the sources. In BSS, the sources have already existed and fixed when they are observed, so (22) is a constant for any perfect solution $\bar{G}$. Finally, one can conclude that the contrast function (4) is maximized if and only if G is a perfect solution. □

3.3. Remarks for the New Proof

The above subsection presents an alternative proof of the contrast function (4) under **P1)**, **P3)** and **A1)**. Assumption **P2)** is not directly required although it was stated by Erdogan [19]. We further analyse **A1)** and **P2)**. Let $\mathbb{V}_{\widehat{B}(S)}$, $\mathbb{V}_S$ and $\mathbb{V}_{S_{i,:}}$ denote the set of vertices of $\widehat{B}(S)$, $Conv(S)$ and $Conv(S_{i,:})$, respectively. For any real-valued vector $S_{i,:}$, $\mathbb{V}_{S_{i,:}} = [\min[S_{i,:}], \max[S_{i,:}]]^T$, where $i \in \{1, \cdots, p\}$. By definition, the set of vertices of $\widehat{B}(S)$ is the Cartesian product of the set of vertices of each source $S_{i,:}$, i.e., $\mathbb{V}_{\widehat{B}(S)} = \mathbb{V}_{S_{1,:}} \otimes \cdots \otimes \mathbb{V}_{S_{p,:}}$. Additionally, **A1)** means that $\mathbb{V}_{\widehat{B}(S)} = \mathbb{V}_S$, so $\mathbb{V}_S =$

$\mathbb{V}_{S_{1,:}} \otimes \cdots \otimes \mathbb{V}_{S_{p,:}}$. On the other hand, it has been shown that **P2)** is equivalent to $ext\ \mathbb{S}_{\check{S}} = ext\ \mathbb{S}_{\check{S}_{1,:}} \otimes \cdots \otimes ext\ \mathbb{S}_{\check{S}_{p,:}}$, where $ext\ \mathbb{S}_{\check{S}_{i,:}}$ denotes the extreme points of the ith source support set, and $ext\ \mathbb{S}_{\check{S}}$ comprises all vertices of the convex hull [17]. The support set is the set of points, for which the probability density function is non-zero. For real-valued signals, the extreme point is equivalent to the vertex. Therefore, **A1)** assumes that in the sample space, the vertices of the bounding hyper-rectangle of S should be observed. In contrast, **P2)** is a theoretical probabilistic assumption, and assumes that the vertices have positive probability. Thus, **A1)** is actually only possible if **P2)** is satisfied. Hence, the three assumptions **P1)**, **P3)** and **A1)** imply the four assumptions **P1)**-**P3)** and **A1)**.

Our proof has shown that elementary matrices and simple factorization of G contribute to prove the contrast function (4). In this subsection, we analyse the imperfect solution by these techniques. In contrast to (15), for any imperfect solution $\underline{\mathrm{G}}$, the simply factorized product must contain at least one E_3 factor. Let F_3 denote the number of E_3 factors, and then we have

$$\underline{G} \triangleq \prod_{f=1}^{F} E_i^f, i \in \{1,2,3\}, F_3 \geqslant 1, \tag{23}$$

where $F = F_1 + F_2 + F_3$. It should be remembered that "$\triangleq$" requires that an identity matrix and an elementary matrix should not be further factorized to two or more elementary matrices. The simply factorized product of G is not unique, and a typical way to simply factorize is factorizing as the steps of Gauss-Jordan elimination.

Remark 1. *Under assumptions **P1)**, **P3)** and **A1)**, we can state that* $vol[\widehat{B}(E_3 S)] > abs(det(E_3))vol[\widehat{B}(S)],\ \forall E_3$.

Proof. Recalling the definition of E_3 in (9), we get

$$E_3 S = E_3 \begin{bmatrix} \cdot \\ S_{u,:} \\ \cdot \\ S_{v,:} \\ \cdot \end{bmatrix} = \begin{bmatrix} \cdot \\ S_{u,:} + kS_{v,:} \\ \cdot \\ S_{v,:} \\ \cdot \end{bmatrix}.$$

From the above equation, E_3 only affects the uth row of S, so we pay attention

to the uth row in the following. Substituting (9) into (6), we get

$$\begin{aligned} \mathrm{vol}[\widehat{B}(E_3S)] &= \prod_{i=1}^{p} \{\max[(E_3S)_{i,:}] - \min[(E_3S)_{i,:}]\} \qquad (24) \\ &= \{\max[(E_3S)_{u,:}] - \min[(E_3S)_{u,:}]\} \prod_{i\neq u} \{\max[S_{i,:}] - \min[S_{i,:}]\}. \end{aligned}$$

Under **A1)** that S contains the vertices of its bounding hyper-rectangle, and as shown in Fig.1, we can be assured the following samples of S (columns of S) exist:

$$\begin{aligned} V_{mm} &= [\ \cdots\ , \max[S_{u,:}],\ \cdots\ , \max[S_{v,:}],\ \cdots\ ]^T \in S, \\ V_{mn} &= [\ \cdots\ , \max[S_{u,:}],\ \cdots\ , \min[S_{v,:}],\ \cdots\ ]^T \in S, \\ V_{nn} &= [\ \cdots\ , \min[S_{u,:}],\ \cdots\ , \min[S_{v,:}],\ \cdots\ ]^T \in S, \\ V_{nm} &= [\ \cdots\ , \min[S_{u,:}],\ \cdots\ , \max[S_{v,:}],\ \cdots\ ]^T \in S. \end{aligned}$$

$$\begin{aligned} \therefore \quad & \max[(E_3S)_{u,:}] - \min[(E_3S)_{u,:}] \qquad (25) \\ = \ & \begin{cases} \max[S_{u,:}] + k\ \max[S_{v,:}] - \min[S_{u,:}] - k\ \min[S_{v,:}], k > 0 \\ \max[S_{u,:}] + k\ \min[S_{v,:}] - \min[S_{u,:}] - k\ \max[S_{v,:}], k < 0 \end{cases} \\ = \ & \max[S_{u,:}] - \min[S_{u,:}] + \mathrm{abs}(k)\{\max[S_{v,:}] - \min[S_{v,:}]\}. \end{aligned}$$

Under **P1)**, $\max[S_{i,:}] - \min[S_{i,:}] > 0,\ i \in \{1, \cdots, p\}$. Additionally, the definition of E_3 implies $k \neq 0$. Hence, (25) leads to

$$\max[(E_3S)_{u,:}] - \min[(E_3S)_{u,:}] > \max[S_{u,:}] - \min[S_{u,:}]. \qquad (26)$$

Substituting (26) into (24), and given $\mathrm{abs}(\det(E_3)) = 1$, we can state more generally that $\mathrm{vol}[\widehat{B}(E_3S)] > \mathrm{abs}(\det(E_3))\mathrm{vol}[\widehat{B}(S)]$. □

Fig.1 illustrates Remark 1 as $p = 2$, for visibility and without loss of generality. As shown on the left side of Fig.1, the region contoured by the bounding hyper-rectangle (the black rectangle) is the same as the convex hull (the green rectangle), which shows that under **A1)** $\widehat{B}(S) = Conv(S)$ and $\mathrm{vol}[\widehat{B}(S)] = \mathrm{vol}[Conv(S)]$. A shear mapping E_3 reshapes the convex hull (the green rectangle) on the left side to another convex hull (the green parallelogram) on the right side, and the volume of the two convex hulls are

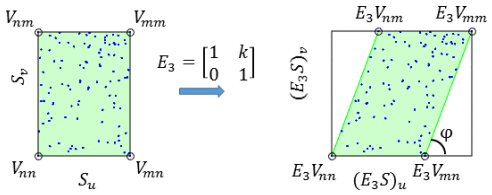

Figure 1: Illustration of Remark 1. The blue dots represent the samples, the regions enclosed by the black rectangles represent the bounding hyper-rectangles, and the green regions represent the convex hulls. On the left side, S contains four vertices expressed by black circles. On the right side, the green parallelogram represents the reshaped bounding hyper-rectangle by E_3, and also is the convex hull of the mapped samples.

the same, i.e., $\text{vol}[Conv(S)] = \text{vol}[Conv(E_3S)]$. Note that on the right side, corresponding edges are tilted with a shear angle φ by E_3, so these orthogonal edges become oblique edges. The shear factor k is the cotangent of φ. Due to the shear angle φ, the maximum and minimum operations of the bounding hyper-rectangle (the black rectangle) expand the parallelogram $Conv(E_3S)$ (the green parallelogram) to a new bounding hyper-rectangle $\widehat{B}(E_3S)$ (the black rectangle). Then, $\text{vol}[\widehat{B}(E_3S)] > \text{vol}[Conv(E_3S)] = \text{vol}[\widehat{B}(S)]$, as shown that the bounding hyper-rectangle (the black rectangle) on the right side is larger than that on the left side.

From a geometric point of view, since the edges of the bounding hyper-rectangle are parallel to the coordinate axes, any edge of the bounding hyper-rectangle is orthogonal to other edges that are connected at the same vertex. The shear mapping E_3 tilts the corresponding edges from orthogonal to oblique, while the reflection mapping E_1 and the scale mapping E_2 preserve the orthogonality of each pair of edges. These geometric features lead to the inequality of Remark 1 and the equalities of (10) and (11). Combining with the simple factorization of G in (15) and (23), the inequality of (21) can be interpreted by the geometric features of elementary matrices.

4. Further Analysis

4.1. Analysis of Assumptions

In this subsection, we discuss the constraints in practical applications required by the four assumptions. **P1)** requires the amplitude of each source varying over a bounded and closed range. Realistic sources randomly vary in time or space over certain ranges, for example audio signals. Therefore, assumption **P1)** can be regarded as weak. In fact, many other BSS algorithms

also require **P1)** implicitly. For instance, ICA algorithms use the higher order statistics (HOS)[3] of the sources, which implies the HOS of the sources is bounded. It follows that the sources are bounded, so ICA algorithms also assume **P1)** implicitly. Assumption **P3)** is a standard BSS assumption. Thus, we turn our attention to the assumptions **A1)** and **P2)**.

In [19], the BCA algorithm was shown to have the potential advantage of performance improvement for short data records, and this advantage was supported by numerical experiments of artificial sources with Copula-t distributions of 10 sources and 5000 samples. The sources with Copula-t distributions are continuous variables, and this is a good example to evaluate the BCA algorithm.

However, we consider the other aspect of the BCA algorithm from the constraint of **A1)**. For short records of continuous variables, **A1)** is a stringent assumption, in particular for tailed distributions[4]. Firstly, the vertex number of the bounding hyper-rectangle grows exponentially with the number of the sources p, that is 2^p. Secondly, **A1)** implies that each edge of the convex hull is orthogonal to other edges connected with the same vertex, as the name "rectangle" suggests. For tailed distributions, the probability of each source attaining the extreme points is low, so the probability of more sources attaining their extreme points at the same sample are much lower.

Let us take audio signals as an example to analyse **A1)**. As shown in Fig.2, for two pairs of audio signals ($p = 2$), the black rectangles represent the bounding hyper-rectangles which are external to the respective convex hulls, and neither the pair of speech signals[5] nor the pair of music solos[6] contains any vertex of their bounding hyper-rectangles. Even if the length of audio sources are increased, **A1)** is still stringent. To demonstrate this, we take one pair of 20-minute Obama talks[7] which are shown on the left side of Fig.3, and one pair of 9-minute music solos[8] which are illustrated on the right side. Similar to Fig.2, neither the pair of speech signals nor the pair

[3]HOS refers to functions which use the higher power of a sample. The third and higher moments, as used in the skewness and kurtosis, are examples of HOS.

[4]Roughly speaking, a tailed distribution is a probability distribution with large skewness or kurtosis, compared to the normal distribution or the exponential distribution.

[5]TIMIT database

[6]https://archive.org/details/solo-piano-7

[7]http://www.obamadownloads.com/obama-mp3s.html

[8]http://www.mfiles.co.uk/classical-mp3.htm

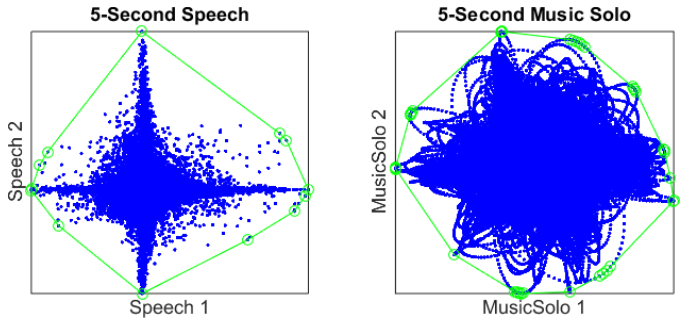

Figure 2: Scatter plot of two pairs of realistic audio signals. Two speech signals sampled at 16 KHz for 5 s are on the left side, while two music solos sampled at 44.1 KHz for 5 s are on the right side. The blue dots represent samples, the black rectangles represent the corresponding bounding hyper-rectangles, and the green lines with green circles represent the corresponding convex hulls.

of music solos contains any vertex of their bounding hyper-rectangles. Since **A1)** requires the sources to contain all of vertices, and the vertex number grows exponentially with the number of the sources, **A1)** can be considered stringent even for long duration of audio sources.

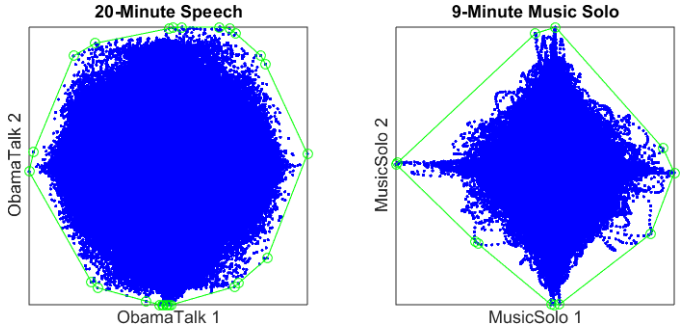

Figure 3: Scatter plot of long audio signals. Two speech signals sampled at 24 KHz for 20 m are on the left side, while two music solos sampled at 44.1 KHz for 9 m are on the right side. The blue dots represent samples, the black rectangles represent the corresponding bounding hyper-rectangles, and the green lines with green circles represent the corresponding convex hulls.

Although the pairs of audio signals are simplistic, these represent the fundamental signals which posed the BSS problem as the so-called "cocktail party problem", and audio signals have been one of the main applications of BSS over the past three decades.

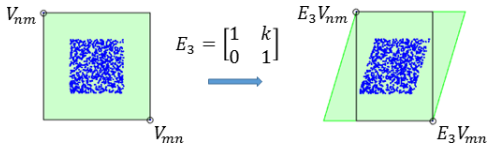

Figure 4: A simple example that E_3 (shear mapping) makes the bounding hyper-rectangle of tailed distributions shrink, given **A1)** is unsatisfied. The blue dots and the black rectangles represent the samples of the corresponding bounding hyper-rectangles, respectively. On the left side, S only contains two vertices V_{mn} and V_{nm}, expressed by black circles. On the right side, E_3V_{mn} and E_3V_{nm} are the respective vertices mapped by E_3 left multiplication, and the green parallelogram is the bounding hyper-rectangle reshaped by E_3.

On the basis of the above discussion, it is reasonable to assume that for BSS problems on tailed distributions, **P1)** and **P3)** are satisfied, but **A1)** is unsatisfied. In such conditions, (10) and (11) still hold, and it follows that (16) still holds. Therefore, even though **A1)** is unsatisfied, under **P1)** and **P3)** the contrast function (4) is preserved for the perfect solution $\bar{G}$. Comparing (15) and (23), we can conclude that the key difference between the perfect solution and the imperfect solution is the E_3 factor.

For E_3 given **A1)** is unsatisfied, Remark 1 does not hold. It follows that there likely exists an E_3 such that $\text{vol}[\widehat{B}(E_3S)] < \text{vol}[\widehat{B}(S)]$. For such a condition, since $\text{abs}(\det(E_3)) = 1$, we get

$$\frac{\text{vol}[\widehat{\varepsilon}(E_3S)]}{\text{vol}[\widehat{B}(E_3S)]} = \frac{\text{abs}(\det(E_3))\text{vol}[\widehat{\varepsilon}(S)]}{\text{vol}[\widehat{B}(E_3S)]} > \frac{\text{vol}[\widehat{\varepsilon}(S)]}{\text{vol}[\widehat{B}(S)]} \tag{27}$$

By comparing (27) and (22), the G that is obtained by maximizing (4) will contain at least one E_3 factor, and thus that G is an imperfect solution. As a result, the BCA algorithm based on (4) will fail, given such E_3 exists.

Given **A1)** is unsatisfied, from a geometric point of view, if there exists an E_3 such that it makes the bounding hyper-rectangle of S shrink, the G obtained by maximizing (4) will be an imperfect solution. For visibility and without loss of generality, we provide a simple two dimensional ($p = 2$) example to demonstrate that for tailed distributions there likely exists an E_3 that makes $\text{vol}[\widehat{B}(S)]$ shrink. Consider two outliers $V_{nm} = [-2c, 2c]^T$, $V_{mn} = [2c, -2c]^T$, $c > 0$, representing the tails of an otherwise uniform distribution in the range $[-c, c]$. It follows that S only contains two out of the four vertices of its bounding hyper-rectangle. As Fig. 4 shows, the E_3 with $k = 0.3$ makes the bounding hyper-rectangle shrink from $16c^2$ to $11.2c^2$.

The existence of such an E_3 is determined by the distribution of sources. For tailed distributions, e.g., audio signals, the performance of BCA contrast function (4) is likely to suffer, since **A1)** is likely to be unsatisfied.

Based on our new proof, we further analyse that the instantaneous BCA algorithm is likely to suffer for tailed distributions, for instance audio signals, because **A1)** is a stringent assumption for continuous sources of tailed distributions. It was briefly stated by Erdogan in conclusion [21] that the performances of the convolutive BCA algorithm is likely to suffer for tailed distributions. Our analysis indicates this limitation is also for the originally instantaneous BCA and is likely to extend to other cases which rely on **A1)**. Although the BCA algorithm is termed as "bounded", the key assumption of BCA in practical applications is **A1)** (or associated **P2)**), rather than the bounded assumption **P1)**.

BCA assumptions	**ICA** assumptions
Bounded **P1)**(explicitly)	Bounded **P1)**(implicitly)
Determined **P3)**	Determined **P3)**
Geometric **A1)**	Stochastic **(Statistical Independence)**

Table 1: Assumptions of BCA and ICA

Although ICA algorithms do not explicitly require that the sources are bounded, ICA algorithms use HOS and calculate HOS based on samples. Hence, ICA algorithms require that HOS of the sources are bounded. It follows that ICA algorithms implicitly assume the sources are bounded. In [19], BCA was considered as a more general approach, covering ICA as a special case for bounded sources. However, we consider that BCA is established based on the geometric assumption **A1)**, while ICA is established based on the stochastic assumption **(Statistical Independence)**, in addition to the bounded assumption **P1)** and the determined assumption **P3)**.

Fundamentally, BCA and ICA are different from the geometric assumption **A1)** and the stochastic assumption **Statistical Independence**, as shown in Table 1. As Fig.5 illustrates, there is an intersection where **A1)** is unsatisfied and independent. Some audio sources can be considered independent but not satisfying **A1)**, which will be shown in the next subsection. Therefore for bounded sources, BCA can be considered as a complementary approach of ICA, rather than a more general approach than ICA.

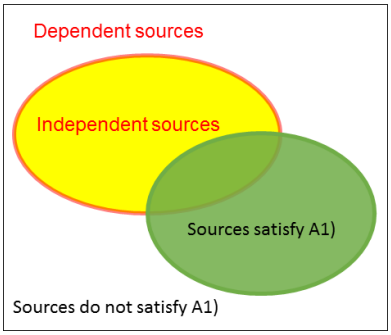

Figure 5: Illustration of the sources divided by the geometric assumption **A1)** and the stochastic assumption **Statistical Independence**. The sources within the green oval satisfy **A1)**, while the sources within the red oval satisfy Statistical Independence.

4.2. Numerical Experiments

In order to investigate our theoretical analysis, we performed simulations on audio sources. We evaluated the BCA algorithm [19] using the MATLAB code[9] provided by Erdogan [19]. Benchmarks were ICA algorithms: FastICA [11], JADE [9] and SOBI [7]. One evaluation criterion is a classical BSS performance index (PI) for the overall mapping $G = WH$ [26] as

$$\begin{aligned} PI(G) &= \frac{1}{2p(p-1)}\left\{\sum_{i=1}^{p}\left(\sum_{j=1}^{p}\frac{|G(i,j)|^2}{\max_k|G(i,k)|^2}-1\right)\right. \\ &\left.+\sum_{j=1}^{p}\left(\sum_{i=1}^{p}\frac{|G(i,j)|^2}{\max_k|G(k,j)|^2}-1\right)\right\}. \end{aligned}$$

The other criterion is the sum square error (SSE) for the separated sources [27] as

$$SSE(S,WY) = \frac{1}{p}\min_{\pi\in\prod_p}\sum_{i=1}^{p}\left\|\frac{S_{i,:}}{\|S_{i,:}\|}-\frac{(WY)_{\pi_i}}{\|(WY)_{\pi_i}\|}\right\|^2,$$

where $\pi = (\pi_1,\ldots,\pi_p)^T$, and $\prod_p = \{\pi \in \mathbb{R}^{p\times 1} \mid \pi_i \neq \pi_j, \forall i \neq j\}$ is the set of all permutations of $\{1,2,\ldots,p\}$. The permutation ambiguity is resolved by the Hungarian algorithm [28] and the authors offer MATLAB

[9] http://aspc.ku.edu.tr/bcasoftware.html

source code [10] online. The scaling ambiguity is normalized by Euclidean norm like $S_{i,:}/\|S_{i,:}\|$. Directly, for the PI and SSE indices the smaller they are, the better the separation quality.

Four experiments were conducted on realistic speech sources and music sources separately. In each experiment, 500 independent runs were performed and averaged. For each independent run, the observed signals Y are artificially generated by randomly selected sources and the mixing matrix H, which was generated randomly from a zero-mean unit-variance normal distribution.

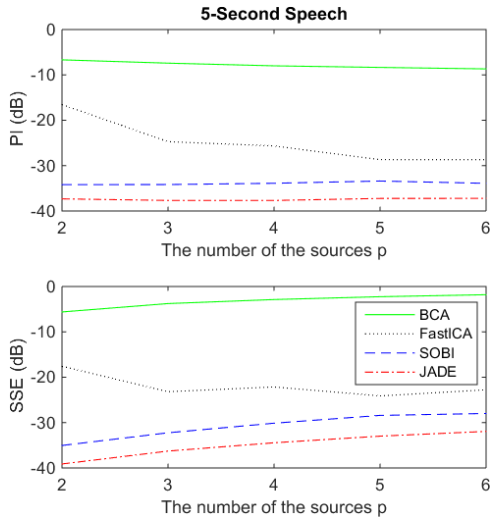

Figure 6: Mean SSE and PI for 5-second speech sources by the number of sources p

The first experiment was conducted on 50 speech segments from the TIMIT dataset, which were sampled at 16 KHz and truncated for 5 seconds. Fig.6 compares, the performance of the three ICA algorithms and BCA algorithm, by the PI and SSE for different number of sources $p = 2, \ldots, 6$. Although the performance of FastICA was worse than the other two ICA algorithms, the three ICA algorithms all achieved satisfactory separation performance. In fact, the determined instantaneous BSS problems on speech sources have been well-developed [2]. However, the separation performance of the BCA algorithm suffered. This result supports our theoretical analysis in the Section 3.1.

[10] http://itakura.ite.tul.cz/zbynek/tddeconv.htm

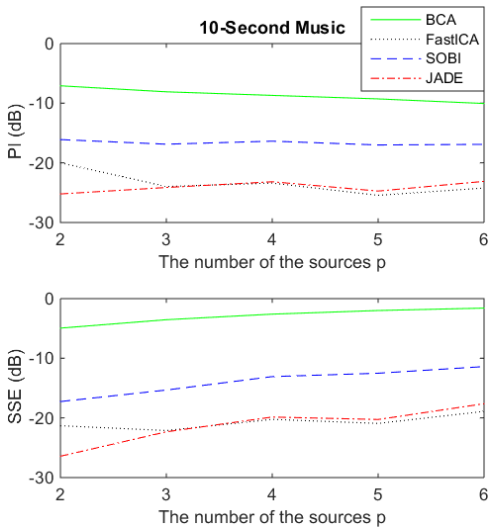

Figure 7: Mean SSE and PI for 10-second music solos by the number of sources p

The second experiment evaluated the PI and SSE on 50 piano[11] and violin[12] segments, which were sampled at 44.1 KHz and truncated for 10 seconds. Fig.7 illustrates the PI and SSE of 10-second music solos separated by the BCA algorithm against the ICA benchmarks for different number of the sources $p = 2, \ldots, 6$. We can also see that the three ICA algorithms outperformed the BCA algorithm on the two criteria, although there were performance gaps between the ICA algorithms. The performance gaps between the BCA algorithm and the worst benchmark are all approximate 10 dB. This result also underpins our theoretical analysis in the above subsection.

Furthermore, to study the separation performance for longer music sources, we conducted the third experiment on the 50 piano and violin segments truncated for 100 seconds. Fig.8 depicts the PI and SSE of 100-second music solos separated by the BCA algorithm against benchmarks for different number of sources $p = 2, \ldots, 6$. We can also see that the three ICA algorithms also outperformed the BCA algorithm on the two criteria. Compared with the separation performance of BCA for 10-second music solos as shown in Fig.7, there was not any substantial improvement for the separation performance

[11]https://archive.org/details/solo-piano-7
[12]http://www.tasminlittle.org.uk

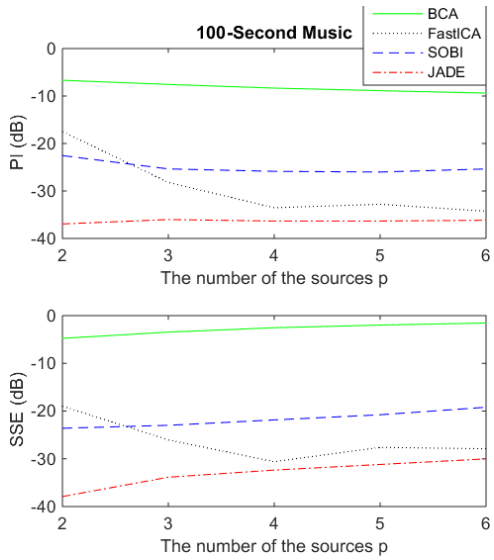

Figure 8: Mean SSE and PI for 100-second music solos by the number of sources p

on tenfold length music solos.

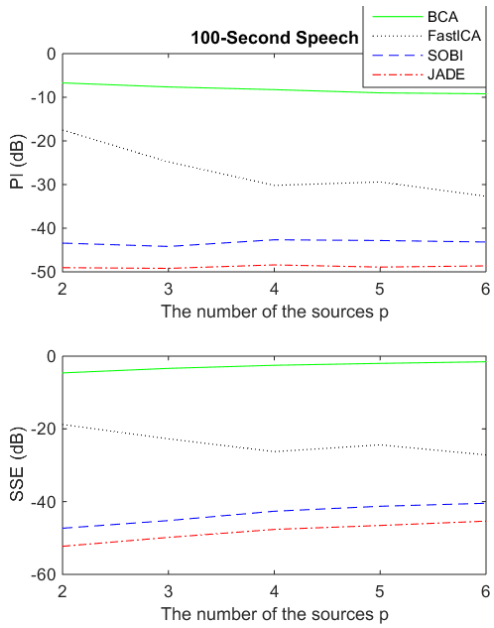

Figure 9: Mean SSE and PI for 100-second speech sources by the number of sources p

Similarly, we conducted the fourth experiment on longer speech sources. 65 speech segments[13] which were sampled at 24 KHz and truncated for 100 seconds. Fig.9 depicts the PI and SSE of 100-second speech sources separated by the BCA algorithm against benchmarks for different number of sources $p = 2, \ldots, 6$. The performance gaps between the BCA algorithm and the worst benchmark are also significant. Compared with the separation performance of BCA for 5-second speech sources as shown in Fig.6, there was not any substantial improvement for the separation performance on twenty-fold length speech sources.

To conclude, the four numerical experiments show that the ICA algorithms outperformed the BCA algorithm. These results support our view that the BCA algorithms based on **A1)** will suffer for continuous tailed-distributed sources, and there are some bounded sources that can be considered independent but not satisfying **A1)**, e.g. audio sources. We have taken audio signals as examples, but the BCA algorithms are also likely to suffer on other tailed distribution signals, such as electroencephalography (EEG) signals, and magnetoencephalography (MEG) signals.

5. Conclusion

This paper presents a new proof of Erdogan's BCA contrast function (4), based on the basic principles of linear algebra and convex geometry. The analysis framework derived from elementary matrices and Gauss-Jordan elimination can be potentially useful in other settings of BSS. Moreover, the link between the key assumptions **A1)** and **P2)** is provided. This paper analyses that **A1)** can be considered as a stringent assumption for continuous sources with tailed distributions, and that **A1)** is stringent in some practical applications, where the performances of the BCA contrast function are likely to suffer. For example, audio sources can be considered are bounded and independent but not satisfying **A1)**. Numerical experiments on audio sources support our analysis. Hence for bounded sources, BCA can be considered as a complementary approach to ICA, not a more general approach than ICA.

[13] http://www.obamadownloads.com/obama-mp3s.html

6. References

[1] A. Hyvarinen, J. Karhunen, E. Oja, Independent component analysis, J. Wiley, New York, 2001.

[2] P. Comon, C. Jutten, Handbook of Blind Source Separation: Independent Component Analysis and Applications, Elsevier Science, Burlington, MA, USA, 2010.

[3] J. F. Cardoso, Blind signal separation: Statistical principles, IEEE Proc. 86 (10) (1998) 2009–2025.

[4] B. Ans, J. Herault, C. Jutten, Adaptive neural architectures: Detection of primitives, in Proc. COGNITIVA (1985) 593–597.

[5] J. Herault, C. Jutten, Space or time adaptive signal process. by neural network models, Intern. Conf. on Neural Networks for Computing (1986) 206–211.

[6] D. T. Pham, Blind separation of instantaneous mixture of sources via an independent component analysis, IEEE Trans. Signal Process. 44 (11) (1996) 2768–2779.

[7] A. Belouchrani, K. AbedMeraim, J. F. Cardoso, E. Moulines, A blind source separation technique using second-order statistics, IEEE Trans. Signal Process. 45 (2) (1997) 434–444.

[8] L. De Lathauwer, B. De Moor, J. Vandewalle, Independent component analysis and (simultaneous) third-order tensor diagonalization, IEEE Trans. Signal Process. 49 (10) (2001) 2262–2271.

[9] J. F. Cardoso, A. Souloumiac, Blind beamforming for non-gaussian signals, IEE Proc. F (Radar and Signal Process.) 140 (6) (1993) 362–370.

[10] J.-F. Cardoso, On the performance of orthogonal source separation algorithms, Proc. EUSIPCO (1994) 776–779.

[11] A. Hyvarinen, Fast and robust fixed-point algorithms for independent component analysis, IEEE Trans. Neural Netw. 10 (3) (1999) 626–634.

[12] D. T. Pham, J. F. Cardoso, Blind separation of instantaneous mixtures of nonstationary sources, IEEE Trans. Signal Process. 49 (9) (2001) 1837–1848.

[13] P. W. Chen, H. Hung, O. Komori, S. Y. Huang, S. Eguchi, Robust independent component analysis via minimum gamma-divergence estimation, IEEE J. Sel. Topics Signal Process. 7 (4) (2013) 614–624.

[14] F. Vrins, J. A. Lee, M. Verleysen, A minimum-range approach to blind extraction of bounded sources, IEEE Trans. Neural Netw. 18 (3) (2007) 809–822.

[15] A. T. Erdogan, A simple geometric blind source separation method for bounded magnitude sources, IEEE Trans. Signal Process. 54 (2) (2006) 438–449.

[16] W. Gao, R. Togneri, V. Sreeram, A contrast function and algorithm for blind separation of audio signals, Proc. Interspeech (2017) 1889–1893.

[17] S. Cruces, Bounded component analysis of linear mixtures: A criterion of minimum convex perimeter, IEEE Trans. Signal Process. 58 (4) (2010) 2141–2154.

[18] S. Cruces, Bounded component analysis of noisy underdetermined and overdetermined mixtures, IEEE Trans. Signal Process. 63 (9) (2015) 2279–2294.

[19] A. T. Erdogan, A class of bounded component analysis algorithms for the separation of both independent and dependent sources, IEEE Trans. Signal Process. 61 (22) (2013) 5730–5743.

[20] H. A. Inan, A. T. Erdogan, A convolutive bounded component analysis framework for potentially nonstationary independent and/or dependent sources, IEEE Trans. Signal Process. 63 (1) (2015) 18–30.

[21] H. A. Inan, A. T. Erdogan, Convolutive bounded component analysis algorithms for independent and dependent source separation, IEEE Trans. Neural Netw. Learn. Syst. 26 (4) (2015) 697–708.

[22] E. Babatas, A. T. Erdogan, Sparse bounded component analysis, Machine Learning for Signal Processing (MLSP), 2016 IEEE 26th International Workshop (2016) 1–6.

[23] H. A. Inan, A. T. Erdogan, S. Cruces, Stationary point characterization for a class of bca algorithms, IEEE Trans. Signal Process. 65 (20) (2017) 5437–5452.

[24] B. Grünbaum, Convex Polytopes, Springer-Verlag, New York, USA, 2003.

[25] P. Gruber, J. Wills, Handbook of Convex Geometry, Elsevier Science, Amsterdam, The Netherlands, 1993.

[26] A. Boudjellal, A. Mesloub, K. Abed-Meraim, A. Belouchrani, Separation of dependent autoregressive sources using joint matrix diagonalization, IEEE Signal Proc. Lett. 22 (8) (2015) 1180–1183.

[27] Z. Y. Yang, Y. Xiang, Y. Rong, K. Xie, A convex geometry-based blind source separation method for separating nonnegative sources, IEEE Trans. Neural Netw. Learn. Syst. 26 (8) (2015) 1635–1644.

[28] P. Tichavsky, Z. Koldovsky, Optimal pairing of signal components separated by blind techniques, IEEE Signal Proc. Lett. 11 (2) (2004) 119–122.