A FAMILY OF BOUNDED COMPONENT ANALYSIS ALGORITHMS

Alper T. Erdogan

Koc University
Electrical and Electronics Engineering Department
Istanbul, Turkey

ABSTRACT

Bounded Component Analysis (BCA) has recently been introduced as an alternative method for the Blind Source Separation problem. Under the generic assumption on source boundedness, BCA provides a flexible framework for the separation of dependent (even correlated) as well as independent sources. This article provides a family of algorithms derived based on the geometric picture implied by the founding assumptions of the BCA approach. We also provide a numerical example demonstrating the ability of the proposed algorithms to separate mixtures of some dependent sources.

Index Terms— Blind Source Separation, Bounded Component Analysis, Independent Component Analysis, Dependent Component Analysis

1. INTRODUCTION

Blind Source Separation is a well-known subject in signal processing (see for example [1] and the references therein). It is the central problem for a diverse set of applications ranging from MIMO communications to brain activity monitoring, and to even non-engineering applications such as financial factor analysis.

Various BSS approaches have been proposed to extract original sources from their mixtures. The hardship caused by the lack of training information and the knowledge about the mixing system is overcome by the additional assumptions made by these approaches. Among these, the assumption about the statistical independence of sources stands out as probably the most popular and the most successful choice. Although the corresponding Independent Component Analysis (ICA) algorithms have been used in various applications, the assumption of independence may not hold depending on the data model.

Recently, in [2], it was shown that if the source distributions have finite support lengths, then the independence assumption can be replaced with a weaker domain separability assumption. This leads to the development of a new BSS approach called Bounded Component Analysis (BCA), which can be used to separate both independent and dependent sources.

The main contribution of this article is to provide a geometric framework for the construction of a family of BCA algorithms. In Section 2, BSS setup assumed in the article and BCA assumptions are introduced. Section 3 is the main part, where the geometric optimization settings for BCA are introduced and equivalence of their global optima to perfect separators are shown. The corresponding iterative algorithms are also provided in the same section. A numerical example to illustrate the dependent component separation performance is given in Section 4. Finally, Section 5 is the conclusion.

2. BCA SETUP

In the article we assume the standard (over)determined instantaneous BSS setup where

- $\mathbf{s} = \begin{bmatrix} s_1 & s_2 & \dots & s_n \end{bmatrix}^T \in \Re^n$ is the zero mean (without loss of generality) and potentially correlated source vector with the covariance $\mathbf{R}_s$. The sources are bounded, i.e. $s_i \in [\alpha_i, \beta_i]$ for all $i = 1, \dots n$ and for some finite $\alpha_i, \beta_i \in \Re$. We also define $\gamma_i = \beta_i - \alpha_i$ as the range of source s_i,
- $\mathbf{H} \in \Re^{m \times n}$ is the mixing matrix (with $m \geq n$),
- $\mathbf{y} = \begin{bmatrix} y_1 & y_2 & \dots & y_m \end{bmatrix}^T \in \Re^m$ is the mixture vector, satisfying $\mathbf{y} = \mathbf{H}\mathbf{s}$, and hence it has the covariance $\mathbf{R}_y = \mathbf{H}\mathbf{R}_s\mathbf{H}^T$.
- $\mathbf{W} \in \Re^{n \times m}$ is the separator matrix to be trained by the BCA Algorithm,
- $\mathbf{z} = \begin{bmatrix} z_1 & z_2 & \dots & z_n \end{bmatrix}^T \in \Re^n$ is the separator output vector, satisfying the equality $\mathbf{z} = \mathbf{W}\mathbf{y}$, and it has the covariance $\mathbf{R}_z = \mathbf{W}\mathbf{R}_y\mathbf{W}^T$,
- We also define $\mathbf{G} = \mathbf{W}\mathbf{H} \in \Re^{p \times p}$ as the overall mapping from sources to separator outputs.

The goal is to train the separator $\mathbf{W}$ based on the available observations of mixtures $\{\mathbf{y}(1), \mathbf{y}(2), \dots, \mathbf{y}(L)\}$.

According to [2], the sufficient conditions for separability are given as

- (A1) $\mathbf{H}$ is full rank,
- (A2) the sources are bounded,
- (A3) the (convex hull of the) domain of the sources can be written as the cartesian product of (the convex hull of) the individual source domains,

Note that the domain separability assumption in (A3) is weaker than the joint pdf separability condition of the independence assumption. In other words, under the generic assumption about source boundedness, BCA provides a more general framework than ICA allowing separation of both independent and dependent sources.

3. A FAMILY OF BCA ALGORITHMS

3.1. A BCA Optimization Setup

We first start by defining the range operator, for individual vector components as

$$\mathcal{R}(z_i) = \max_{k\in\{1\ldots L\}} z_i(k) - \min_{k\in\{1\ldots L\}} z_i(k), \tag{1}$$

and for vectors as $\mathcal{R}(\mathbf{z}) = \begin{bmatrix} \mathcal{R}(z_1) & \ldots & \mathcal{R}(z_n) \end{bmatrix}^T$, i.e., the vector composed of individual component ranges.

The family of BCA algorithms are defined by the optimization settings of the form

$$\text{maximize} \qquad J(\mathbf{W}) = \frac{\sqrt{\det(\mathbf{R}_z)}}{f(\mathcal{R}(\mathbf{z}))}. \tag{2}$$

In the above formulation we consider two geometric objects corresponding to vectors (of sources, mixtures or outputs), whose illustrations are provided in Figure 1 for the numerical example in Section 4:

- **Bounding Hyperrectangle** corresponds to the box defined by the Cartesian product of the support sets of the individual components. This can be also defined as the minimum volume box containing all samples and aligning with the coordinate axes.
- **Principal Hyperellipse** is the hyperellipse whose principal semi-axis directions are determined by the eigenvectors of the covariance matrix and whose principal semi-axis lengths are equal to principal standard deviations, i.e., the square roots of the eigenvalues of the covariance matrix.

The numerator of the objective function is the (scaled) volume of the principal hyperellipse, whereas the denominator is a measure of the bounding hyperrectangle for the output vectors.

We can propose different alternatives for the boundary measure function f:

- An intuitive choice for f would be the mapping defined by taking product of the elements of its argument, i.e., we can define

$$f_1(\mathcal{R}(\mathbf{z})) = \mathcal{R}(z_1)\mathcal{R}(z_2)\ldots\mathcal{R}(z_n), \tag{3}$$

which corresponds to the volume of the bounding hyperrectangle. Therefore, the optimization in (2) with f_1 in (3) corresponds to maximization of the volume of the principal hyperellipse corresponding to output samples relative to the volume of the bounding box.

It is interesting to note that the logarithm of the objective function in (2) with f_1 is equal to

$$\log(J_1(\mathbf{W})) = \frac{1}{2}\log\det(\mathbf{R}_z) - \sum_{i=1}^{n}\log(\mathcal{R}(z_i)), \tag{4}$$

which is the Pham's approximation in [3] for the Mutual Information based objective function obtained by the use of quantiles. As will be shown in the next section, under the BCA assumptions, the maximization of this objective function leads to separation condition. Therefore, the objective function proposed by Pham in ICA framework is also a valid objective function for BCA.

- Another interesting choice would be defining

$$f_2(\mathcal{R}(\mathbf{z})) = \|\mathcal{R}(\mathbf{z})\|^n \tag{5}$$

where $\|.\|$ is a norm defined over the vector space $\Re^n$. This choice corresponds to the main diagonal length of the bounding hyperrectangle. The choice of the corresponding norm would result in different objective functions in (2). We can pick l_p norms as illustrative examples.

3.2. The Global Optimality of the Perfect Separators

In this section, we show that the global optima of the optimization problems in (2) correspond to some perfect separators.

We first start by noting that, under the BCA's domain separability assumption, we can write

$$\mathcal{R}(z_i) = \|\mathbf{G}_{i,:}\mathbf{\Gamma}_s\|_1, \tag{6}$$

where $\mathbf{G}_{i,:}$ is the i^{th} row of $\mathbf{G}$ and $\mathbf{\Gamma}_s = diag(\gamma_1, \gamma_2, \ldots, \gamma_p)$ is the diagonal matrix containing range values for the sources. We can further define $\mathbf{C} = \mathbf{G}\mathbf{\Gamma}_s$, in which case we can write

$$\mathcal{R}(z_i) = \|\mathbf{C}_{i,:}\|_1. \tag{7}$$

Therefore, we can write the corresponding objective function more explicitly in terms of $\mathbf{C}$ as

$$J(\mathbf{C}) = \frac{\sqrt{det(\mathbf{R}_s)}}{det(\mathbf{\Gamma}_s)} \frac{|\det(\mathbf{C})|}{f\left(\begin{bmatrix} \|\mathbf{C}_{1,:}\|_1 & \ldots & \|\mathbf{C}_{n,:}\|_1 \end{bmatrix}^T\right)} \tag{8}$$

3.2.1. Case 1: f_1 in (3)

The objective function in this case is equivalent to

$$J_1(\mathbf{C}) = \frac{\sqrt{det(\mathbf{R}_s)}}{det(\mathbf{\Gamma}_s)} \frac{|\det(\mathbf{C})|}{\prod_{i=1}^{n} \|\mathbf{C}_{i,:}\|_1}. \tag{9}$$

The equivalence of the global optima to some perfect separators follows from

$$\begin{aligned} |\det(\mathbf{C})| &\leq \prod_{i=1}^{n} \|\mathbf{C}_{i,:}\|_2 & (10) \\ &\leq \prod_{i=1}^{n} \|\mathbf{C}_{i,:}\|_1, & (11) \end{aligned}$$

where the inequality in (10) is the Hadamard inequality and the inequality in (11) is due to ordering between p-norms. The equality in (10) is achieved iff the rows of $\mathbf{C}$ are orthogonal and the equality in (11) is achieved iff the rows of $\mathbf{C}$ align with coordinate axes. Therefore, the equality holds if and only if $\mathbf{C} = \mathbf{PD}$, where $\mathbf{P}$ is a permutation matrix and $\mathbf{D}$ is a nonsingular diagonal matrix.

3.2.2. Case 2: f_2 in (5)

We perform analysis for particular l_p norms used in f_2:

- l_1 norm based f_2: In this case the objective function is equal to

$$J_{2,1}(\mathbf{C}) = \frac{\sqrt{det(\mathbf{R}_s)}}{det(\mathbf{\Gamma}_s)} \frac{|\det(\mathbf{C})|}{(\sum_{i=1}^{n} \|\mathbf{C}_{i,:}\|_1)^n}. \tag{12}$$

  We note that

$$\begin{aligned} |\det(\mathbf{C})| &\leq (\frac{1}{n}\sum_{i=1}^{n} \|\mathbf{C}_{i,:}\|_1)^n & (13) \\ &= \frac{1}{n^n}(\sum_{i=1}^{n} \|\mathbf{C}_{i,:}\|_1)^n & (14) \end{aligned}$$

  where the inequality in (13) is obtained by applying arithmetic-geometric mean inequality (AGMI) to (11). Note that the equality in the AGMI holds iff all rows of $\mathbf{C}$ have the same 1-norm. Therefore, for the objective function, we can write

$$J_{2,1}(\mathbf{C}) \leq \frac{\sqrt{det(\mathbf{R}_s)}}{det(\mathbf{\Gamma}_s)} \frac{1}{n^n} \tag{15}$$

  where the bound is achieved only for the perfect separators of the form

$$\mathbf{C} = d\mathbf{P}\mathbf{\Upsilon}, \tag{16}$$

  where d is an arbitrary non-negative scaling, $\mathbf{P}$ is a permutation matrix and $\mathbf{\Upsilon}$ is a diagonal matrix with ±1's on the diagonal.

- l_2 norm based f_2: The corresponding objective function is

$$J_{2,2}(\mathbf{C}) = \frac{\sqrt{det(\mathbf{R}_s)}}{det(\mathbf{\Gamma}_s)} \frac{|\det(\mathbf{C})|}{(\sum_{i=1}^{n} \|\mathbf{C}_{i,:}\|_1^2)^{n/2}}. \tag{17}$$

  Using the fact $\|\mathcal{R}(\mathbf{z})\|_1 \leq \sqrt{n}\|\mathcal{R}(\mathbf{z})\|_2$ and the upper bound in (15), we can obtain the upper bound expression for $J_{2,2}$ as

$$J_{2,2}(\mathbf{C}) \leq \frac{\sqrt{det(\mathbf{R}_s)}}{det(\mathbf{\Gamma}_s)} \frac{1}{n^{n/2}} \tag{18}$$

  where the bound is achieved only for the perfect separators in (16).

- l_∞ norm based f_2: The corresponding objective function is

$$J_{2,\infty}(\mathbf{C}) = \frac{\sqrt{det(\mathbf{R}_s)}}{det(\mathbf{\Gamma}_s)} \frac{|\det(\mathbf{C})|}{(\max_{i=1,\ldots,n} \|\mathbf{C}_{i,:}\|_1)}. \tag{19}$$

  Using the norm inequality, $\|\mathcal{R}(\mathbf{z})\|_1 \leq n\|\mathcal{R}(\mathbf{z})\|_\infty$ and the upper bound in (15), the upper bound expression for $J_{2,\infty}$ can be written as

$$J_{2,\infty}(\mathbf{C}) \leq \frac{\sqrt{det(\mathbf{R}_s)}}{det(\mathbf{\Gamma}_s)}. \tag{20}$$

  It is achieved only for the perfect separators in (16).

3.3. Adaptive Algorithms

In this section, we provide the adaptive algorithms corresponding to the optimization setting in (2). We first note that the range operator defined in (1) is a non-differentiable convex function of $\mathbf{W}$. The subdifferential set for this function at a given point $\mathbf{W}$ is given by

$$\partial\mathcal{R}(z_i) = \{ \sum_{k\in\mathcal{I}_{i,max}} \lambda_k \mathbf{e}_i \mathbf{y}(k)^T - \sum_{l\in\mathcal{I}_{i,min}} \eta_l \mathbf{e}_i \mathbf{y}(l)^T \mid \lambda_k \geq 0, \sum_{k\in\mathcal{I}_{i,max}} \lambda_k = 1, \eta_l \geq 0, \sum_{l\in\mathcal{I}_{i,min}} \eta_l = 1\}, \tag{21}$$

where $\mathcal{I}_{i,max}$ is the set of indexes where maximum for z_i is achieved, $\mathcal{I}_{i,min}$ is the set of indexes where minimum for z_i is achieved and $\mathbf{e}_i$ is the standard basis vector. Note that the subdifferential set is constructed from the convex hull of rank one matrices whose i^{th} row contains an input vector causing either maximum or minimum output, and other rows are zero.

The subgradient based iterative algorithms for maximizing objective functions introduced in Section 3.1 is provided below:

- Iterative algorithm for f_1: The algorithm iterations to maximize $\log(J_1(\mathbf{W}))$ can be written as

$$\mathbf{W}^{(j+1)} = \mathbf{W}^{(j)} + \mu^{(j)}((\mathbf{W}^{(j)}\mathbf{R}_y\mathbf{W}^{(j)^T})^{-1}\mathbf{W}^{(j)}\mathbf{R}_\mathbf{y} - \sum_{i=1}^{n} \frac{1}{\mathcal{R}(z_i^{(j)})}\mathbf{e}_i(\mathbf{y}(k_{i,max}^{(j)}) - \mathbf{y}(k_{i,min}^{(j)}))^T), \tag{22}$$

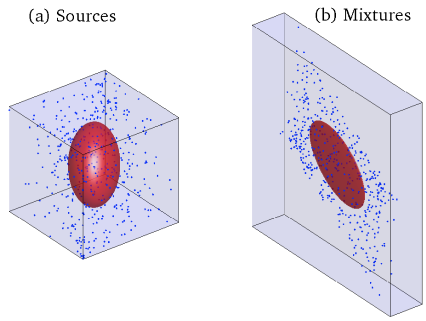

Fig. 1. Samples, Bounding Hyperrectangle and Principal Hyperellipse for (a) Sources (b) Mixtures ($\rho = 0.5$).

where

- $\mu^{(j)}$ is the step size at the j^{th} iteration,
- $k_{i,max}^{(i)}$ is an index from the set $\mathcal{I}_{i,max}^{(j)}$, i.e., a sample index where maximum at the i^{th} component is achieved at the j^{th} iteration,
- $k_{i,min}^{(j)} \in \mathcal{I}_{i,min}^{(j)}$ is a sample index for the minimum at the j^{th} iteration.

- Iterative algorithm for f_2: The algorithm iterations to maximize $\log(J_{2,p}(\mathbf{W}))$ can be written as, for p=1,2,

$$\mathbf{W}^{(j+1)} = \mathbf{W}^{(j)} + \mu^{(j)}((\mathbf{W}^{(j)}\mathbf{R}_y {\mathbf{W}^{(j)}}^T)^{-1}\mathbf{W}^{(j)}\mathbf{R}_\mathbf{y} - \frac{n}{\|\mathcal{R}(\mathbf{z}^{(j)})\|_p}\sum_{i=1}^{n}\mathbf{e}_i(|z_i^{(j)}(k_{i,max})|^{p-1}\mathbf{y}(k_{i,max}^{(j)}) - |z_i^{(j)}(k_{i,min}^{(j)})|^{p-1}\mathbf{y}(k_{i,min}^{(j)}))^T), \quad (23)$$

and, for $p = \infty$ as

$$\mathbf{W}^{(j+1)} = \mathbf{W}^{(j)} + \mu^{(j)}((\mathbf{W}^{(j)}\mathbf{R}_y {\mathbf{W}^{(j)}}^T)^{-1}\mathbf{W}^{(j)}\mathbf{R}_\mathbf{y} - \frac{n}{\|\mathcal{R}(\mathbf{z}^{(j)})\|_\infty}\sum_{i\in\mathcal{I}_{\mathcal{R}(\mathbf{z}^{(j)})}}\zeta_i\mathbf{e}_i(\mathbf{y}(k_{i,max}^{(j)}) - \mathbf{y}(k_{i,min}^{(j)}))^T),$$

where

- $\mathcal{I}_{\mathcal{R}(\mathbf{z}^{(\mathbf{j})})} = \{i \mid \mathcal{R}(z_i^{(j)}) = \|\mathcal{R}(\mathbf{z}^{(j)})\|_\infty\}$, i.e., the set of output indexes for which the maximum range is achieved at the j^{th} iteration,
- ζ_i's are the convex combination weights, which have the property $\zeta_i \geq 0, \forall i \in \mathcal{I}_{\mathcal{R}(\mathbf{z}^{(\mathbf{j})})}$ and $\sum_{i\in\mathcal{I}_{\mathcal{R}(\mathbf{z}^{(j)})}} \zeta_i = 1$.

4. NUMERICAL EXAMPLE

We consider a numerical example demonstrating dependent source separation capability of the proposed BCA algorithms. We consider a BSS scenario with 3 sources and 3 mixtures (to enable 3D pictures of samples) for a randomly selected channel. The sources have the Copula-t distribution with 4 degrees of freedom and with a Toeplitz correlation parameter matrix whose first row is $[\ 1 \quad \rho \quad \rho^2\ ]$. In Figure 1, the illustrating pictures for samples, bounding hyperrectangle and principal hyperellipse are provided for both sources and mixtures.

Figure 2 shows the output total signal energy to total interference energy (over all outputs) ratio for BCA -$J_{2,1}$ and FastICA algorithms and for different mixture block lengths. As the dependent sources violates ICA assumptions, FastICA's perfromance degrades with source correlation. The degradation in BCA algorithm performance for large ρ values is due to inadequate representation of the "source domain corner points" with finite data, which can be seen by the improvement due to increase in block size length.

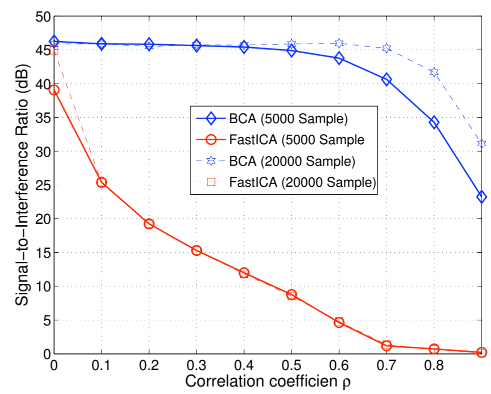

Fig. 2. Dependent source separation performance results.

5. CONCLUSION

This article provided a family of BCA algorithms based on a special geometric optimization setup whose global optima are shown to be perfect separators. These algorithms can be used for BSS problems involving both independent and dependent signals, for the signal models obeying BCA assumptions.

6. REFERENCES

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

[2] 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.

[3] Dinh-Tuan 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.