AN EXTENDED FAMILY OF BOUNDED COMPONENT ANALYSIS ALGORITHMS

Huseyin A. Inan, and Alper T. Erdogan

Electrical and Electronics Engineering Department,
Koc University,
Istanbul, Turkey

ABSTRACT

Bounded Component Analysis (BCA) is a recent concept proposed as an alternative method for Blind Source Separation problem. BCA enables the separation of dependent as well as independent sources from their mixtures under the practical assumption on source boundedness. This article extends the optimization setting of a recent BCA approach which can be used to produce a variety of BCA algorithms. The article also provides examples of objective functions produced from the proposed optimization setting and the corresponding iterative algorithms. The numerical examples illustrate the advantages of proposed BCA examples regarding the correlated source separation capability over the state of the art ICA based approaches.

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

1. INTRODUCTION

Bounded Component Analysis (BCA) is a new Blind Source Separation (BSS) approach introduced in [1] that can separate both independent and dependent sources, providing a more general framework than ICA when the sources are known to be bounded.

In this article, we extend the BCA approach considered in [2] by providing a general optimization framework which can be used to produce numerous BCA algorithms. The approach in [2] exploits two geometric objects defined on output samples which are principal hyper-ellipsoid and bounding hyper-rectangle and optimize the relative sizes of these objects where the volume and the main diagonal length is considered to determine the size of bounding hyper-rectangle. In this article, we generalize the approach in [2] by considering more general functions of ranges of output samples (corresponding to the side lengths of bounding hyper-rectangle) which also covers the size of bounding hyper-rectangle. In addition, we provide some examples of these functions which produce a variety of BCA algorithms.

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

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

2. EXTENDED BCA APPROACH

We consider the following BSS setup in the article:

- We have a deterministic setup consisting of p real sources $\boldsymbol{s} = \begin{bmatrix} s_1 & s_2 & \dots & s_p \end{bmatrix}^T$. We assume that the sources are bounded $s_m(k) \in [\alpha_m, \beta_m]$ for $m = 1, ..., p$ and $k \in \mathbb{Z}$.

  In this article, we do not assume that the sources are independent, or uncorrelated. In fact, the sources are allowed to be potentially correlated.

- The sources are mixed by a linear and instantaneous system $\boldsymbol{H} \in \Re^{q \times p}$ where we assume $q \geq p$. The mixtures $\boldsymbol{y} = \begin{bmatrix} y_1 & y_2 & \dots & y_q \end{bmatrix}^T$ and the sources are related by $\boldsymbol{y} = \boldsymbol{Hs}$.

- $\boldsymbol{W} \in \Re^{p \times q}$ is the separator matrix of the system which produces the outputs as $\mathbf{o} = \boldsymbol{Wy}$.

- The overall system function is defined as $\boldsymbol{G} = \boldsymbol{WH} \in \Re^{p \times p}$, therefore, it satisfies $\mathbf{o} = \boldsymbol{Gs}$.

There is a finite set of observations of mixture samples $Y = \{\boldsymbol{y}(1), \boldsymbol{y}(2), \dots, \boldsymbol{y}(N)\}$ with the corresponding set of unobservable source samples $S = \{\boldsymbol{s}(1), \boldsymbol{s}(2), \dots, \boldsymbol{s}(N)\}$. The set of output samples are produced as

$$\begin{aligned} O &= \{\boldsymbol{Wy}(1), \boldsymbol{Wy}(2), \dots, \boldsymbol{Wy}(N)\} \\ &= \{\boldsymbol{Gs}(1), \boldsymbol{Gs}(2), \dots, \boldsymbol{Gs}(N)\}. \end{aligned}$$

We introduce the following assumption regarding the set S:
Assumption: S contains the vertices of its (non-degenerate) bounding hyper-rectangle (**A1**).

We provide the extended BCA optimization framework by modifying the denominator of the objective functions of [2] by considering generalized functions of ranges of outputs as

$$\boxed{J(\boldsymbol{W})=\frac{\sqrt{det(\hat{\boldsymbol{R}}_{\mathbf{o}})}}{f\left(\hat{\mathcal{R}}(o_1),\hat{\mathcal{R}}(o_2),\ldots,\hat{\mathcal{R}}(o_p)\right)},} \tag{1}$$

where $\hat{\boldsymbol{R}}_{\mathbf{o}}=\frac{1}{N}\sum_{l=1}^{N}(\mathbf{o}(l)-\hat{\boldsymbol{\mu}}(\mathbf{o}))(\mathbf{o}(l)-\hat{\boldsymbol{\mu}}(\mathbf{o}))^T$, $\hat{\boldsymbol{\mu}}(\mathbf{o})=\frac{1}{N}\sum_{l=1}^{N}\mathbf{o}(l)$ is the sample covariance matrix of $\mathbf{o}$, $\hat{\mathcal{R}}(o_m)$ is the range of the m'th component of the vectors in the set O and f is any function that satisfies the following:

$$f\left(\hat{\mathcal{R}}(o_1),\hat{\mathcal{R}}(o_2),\ldots,\hat{\mathcal{R}}(o_p)\right)\geq c_p\prod_{m=1}^{p}\hat{\mathcal{R}}(o_m), \tag{2}$$

where the equality is achievable for a finite constant c_p with some specific requirements.

Theorem: Assuming $\boldsymbol{H}$ is a full rank matrix and $\hat{\boldsymbol{R}}_{\boldsymbol{S}}\succ 0$, the set of global maxima for J in (1) is equal to a set of perfect separator matrices.

Proof: We first note that $\hat{\boldsymbol{R}}_{\mathbf{o}}=\boldsymbol{G}\hat{\boldsymbol{R}}_{\boldsymbol{S}}\boldsymbol{G}^T$ which yields $\sqrt{det(\hat{\boldsymbol{R}}_{\mathbf{o}})}=|det(\boldsymbol{G})|\sqrt{det(\hat{\boldsymbol{R}}_{\boldsymbol{S}})}$.

When assumption **(A1)** holds, we can write $\hat{\mathcal{R}}(o_m)=||\boldsymbol{G}_{m,:}\boldsymbol{\Upsilon}||_1$ where $\boldsymbol{G}_{m,:}$ is the m^{th} row of $\boldsymbol{G}$ and $\boldsymbol{\Upsilon}=diag(\hat{\mathcal{R}}(s_1),\hat{\mathcal{R}}(s_2),\ldots,\hat{\mathcal{R}}(s_p))$. We define $\boldsymbol{A}=\boldsymbol{G}\boldsymbol{\Upsilon}$ and write (1) in terms of $\boldsymbol{A}$ as

$$\begin{aligned}J(\boldsymbol{W})&=\frac{|det(\boldsymbol{A}\boldsymbol{\Upsilon}^{-1})|\sqrt{det(\hat{\boldsymbol{R}}_{\boldsymbol{S}})}}{f\left(||\boldsymbol{A}_{1,:}||_1,||\boldsymbol{A}_{2,:}||_1,\ldots,||\boldsymbol{A}_{p,:}||_1\right)},\\&=\frac{\sqrt{det(\hat{\boldsymbol{R}}_{\boldsymbol{S}})}}{\prod_{m=1}^{p}\hat{\mathcal{R}}(s_m)}\frac{|det(\boldsymbol{A})|}{f\left(||\boldsymbol{A}_{1,:}||_1,||\boldsymbol{A}_{2,:}||_1,\ldots,||\boldsymbol{A}_{p,:}||_1\right)}\end{aligned}$$

Using the Hadamard inequality [3] and the ordering $||\boldsymbol{q}||_1\geq||\boldsymbol{q}||_2$ for any $\boldsymbol{q}$ yields

$$det(\boldsymbol{A})\leq\prod_{m=1}^{p}||\boldsymbol{A}_{m,:}||_2 \tag{3}$$

$$\leq\prod_{m=1}^{p}||\boldsymbol{A}_{m,:}||_1. \tag{4}$$

Since the function f satisfies (2), we obtain

$$\frac{|det(\boldsymbol{A})|}{f\left(||\boldsymbol{A}_{1,:}||_1,||\boldsymbol{A}_{2,:}||_1,\ldots,||\boldsymbol{A}_{p,:}||_1\right)}\leq\frac{\prod_{m=1}^{p}||\boldsymbol{A}_{m,:}||_1}{c_p\prod_{m=1}^{p}||\boldsymbol{A}_{m,:}||_1},$$

which further implies

$$J(\boldsymbol{W})\leq\frac{1}{c_p}\frac{\sqrt{det(\hat{\boldsymbol{R}}_{\boldsymbol{S}})}}{\prod_{m=1}^{p}\hat{\mathcal{R}}(s_m)}. \tag{5}$$

To achieve the equality in (5), the equalities in (2), (3) and (4) must be achieved. The equality in (3) is achieved if and only if the rows of $\boldsymbol{A}$ are orthogonal to each other and the equality in (4) is achieved if and only if the rows of $\boldsymbol{A}$ align with the coordinate axes. Therefore, the equality in (3) and (4) holds if and only if $\boldsymbol{A}=\boldsymbol{P}\boldsymbol{D}$, hence, $\boldsymbol{G}=\boldsymbol{P}\boldsymbol{D}\boldsymbol{\Upsilon}^{-1}$ where $\boldsymbol{P}$ is a permutation matrix and $\boldsymbol{D}$ is a nonsingular diagonal matrix which corresponds to the perfect separators. Hence, with the specific requirement of (2), the global maxima of the objective function (1) correspond to the perfect separators or a subset of perfect separators.

We here provide some examples for the function f:

- $f_1\left(\hat{\mathcal{R}}(o_1),\hat{\mathcal{R}}(o_2),\ldots,\hat{\mathcal{R}}(o_p)\right)=\prod_{m=1}^{p}\hat{\mathcal{R}}(o_m)$: This is a trivial example for which the global maxima is achieved when $\boldsymbol{G}=\boldsymbol{P}\boldsymbol{D}\boldsymbol{\Upsilon}^{-1}$ corresponding to perfect separators. We note that this function corresponds to the volume of the bounding hyperrectangle of outputs and is equivalent to the objective function $J_1^{(\boldsymbol{W})}(\boldsymbol{W})$ of [2].

- $f_{2,r}\left(\hat{\mathcal{R}}(o_1),\hat{\mathcal{R}}(o_2),\ldots,\hat{\mathcal{R}}(o_p)\right)=\left\|\begin{bmatrix}\hat{\mathcal{R}}(o_1) & \hat{\mathcal{R}}(o_2) & \ldots & \hat{\mathcal{R}}(o_p)\end{bmatrix}^T\right\|_r^p$ where $r\geq 1$: From the ordering $||\boldsymbol{q}||_r\geq p^{\frac{1-r}{r}}||\boldsymbol{q}||_1$ for any $\boldsymbol{q}\in\Re^p$ and the Arithmetic-Geometric-Mean-Inequality, we have

$$\begin{aligned}&f_{2,r}\left(\hat{\mathcal{R}}(o_1),\hat{\mathcal{R}}(o_2),\ldots,\hat{\mathcal{R}}(o_p)\right)\\&\geq p^{\frac{p(1-r)}{r}}\left\|\begin{bmatrix}\hat{\mathcal{R}}(o_1) & \hat{\mathcal{R}}(o_2) & \ldots & \hat{\mathcal{R}}(o_p)\end{bmatrix}^T\right\|_1^p\\&\geq p^{\frac{p}{r}}\prod_{m=1}^{p}\hat{\mathcal{R}}(o_m).\end{aligned}$$

  The equality is achieved when $c_p=p^{\frac{p}{r}}$ and the ranges $\hat{\mathcal{R}}(o_m)$ for $m=1,2,\ldots p$ are equal to each other. Hence, the global maxima is achieved when $\boldsymbol{G}=k\text{diag}(\rho)\boldsymbol{P}\boldsymbol{\Upsilon}^{-1}$ where $\rho\in\{-1,1\}^p$ corresponding to a subset of perfect separators. We note that these functions correspond to the length of the main diagonal of the bounding hyperrectangle of outputs and are equivalent to the objective functions $J_{2,r}^{(\boldsymbol{W})}(\boldsymbol{W})$ of [2] for $r=1,2,\infty$.

- $f_3\left(\hat{\mathcal{R}}(o_1),\hat{\mathcal{R}}(o_2),\hat{\mathcal{R}}(o_3)\right)=\left(2\hat{\mathcal{R}}(o_1)\hat{\mathcal{R}}(o_2)+2\hat{\mathcal{R}}(o_1)\hat{\mathcal{R}}(o_3)+2\hat{\mathcal{R}}(o_2)\hat{\mathcal{R}}(o_3)\right)^{\frac{3}{2}}$: This example illustrates the surface area of the bounding hyperrectange of outputs for $p=3$. Using the Arithmetic-Geometric-Mean-Inequality (AGMI) yields

$$f_3\left(\hat{\mathcal{R}}(o_1),\hat{\mathcal{R}}(o_2),\hat{\mathcal{R}}(o_3)\right)\geq 6^{3/2}\prod_{m=1}^{3}\hat{\mathcal{R}}(o_m),$$

where the equality is achieved in the same condition with the norm example for $c_p = 6^{3/2}$. We can generalize this by choosing $f_3\left(\hat{\mathcal{R}}(o_1), \hat{\mathcal{R}}(o_2), \ldots, \hat{\mathcal{R}}(o_p)\right) = \left(\sum_{t=1}^{p} \hat{\mathcal{R}}(o_1)^{m_{t,1}} \hat{\mathcal{R}}(o_2)^{m_{t,2}} \ldots \hat{\mathcal{R}}(o_p)^{m_{t,p}}\right)^{p/x}$ where $\sum_{t=1}^{p} m_{t,j} = x$ for $j = 1, 2, \ldots p$ and $x \in \Re^+$. The global maxima correspond to a subset of perfect separators.

- $f_4\left(\hat{\mathcal{R}}(o_1), \hat{\mathcal{R}}(o_2), \ldots, \hat{\mathcal{R}}(o_p)\right) = log\left(e^{\hat{\mathcal{R}}(o_1)} + e^{\hat{\mathcal{R}}(o_2)} + \ldots + e^{\hat{\mathcal{R}}(o_p)}\right)^{2p}$:
In this case, using the AGMI yields

$$\begin{aligned}
& f_4\left(\hat{\mathcal{R}}(o_1), \hat{\mathcal{R}}(o_2), \ldots, \hat{\mathcal{R}}(o_p)\right) \\
& \geq log\left(p\left(e^{\hat{\mathcal{R}}(o_1)+\hat{\mathcal{R}}(o_2)+\ldots+\hat{\mathcal{R}}(o_p)}\right)^{1/p}\right)^{2p} \\
& \geq \left(log(p) + \left(\hat{\mathcal{R}}(o_1)\hat{\mathcal{R}}(o_2)\ldots\hat{\mathcal{R}}(o_p)\right)^{1/p}\right)^{2p} \\
& \geq 2^{2p} log(p)^p \prod_{m=1}^{p} \hat{\mathcal{R}}(o_m).
\end{aligned}$$

The equality is achieved when $c_p = 2^{2p} log(p)^p$ and the ranges $\hat{\mathcal{R}}(o_m)$ for $m = 1, 2, \ldots p$ are equal to each other and $log(p) = \left(\hat{\mathcal{R}}(o_1)\hat{\mathcal{R}}(o_2)\ldots\hat{\mathcal{R}}(o_p)\right)^{1/p}$, yielding $\hat{\mathcal{R}}(o_m) = log(p)$ for $m = 1, 2, \ldots p$. Hence, the global maxima is achieved when $\boldsymbol{G} = log(p)\boldsymbol{P}\boldsymbol{\Upsilon}^{-1}$ corresponding to a subset of perfect separators.

For the corresponding iterative algorithms, rather than maximizing J , we maximize its logarithm to simplify the update components:

$$\begin{aligned}
\bar{J}(\boldsymbol{W}) &= log\left(J(\boldsymbol{W})\right) = \frac{1}{2} log\left(det\left(\boldsymbol{W}\hat{\boldsymbol{R}}_{\boldsymbol{y}}\boldsymbol{W}^T\right)\right) \\
&- log\left(f\left(\hat{\mathcal{R}}(o_1), \hat{\mathcal{R}}(o_2), \ldots, \hat{\mathcal{R}}(o_p)\right)\right).
\end{aligned}$$

The derivative of the first part of $\bar{J}(\boldsymbol{W})$ with respect to $\boldsymbol{W}$ is

$$\frac{1}{2}\frac{\partial log\left(det\left(\boldsymbol{W}\hat{\boldsymbol{R}}_{\boldsymbol{y}}\boldsymbol{W}^T\right)\right)}{\partial \boldsymbol{W}} = \left(\boldsymbol{W}\hat{\boldsymbol{R}}_{\boldsymbol{y}}\boldsymbol{W}^T\right)^{-1}\boldsymbol{W}\hat{\boldsymbol{R}}_{\boldsymbol{y}}.$$

We note that since f_1 and $f_{2,r}$ functions for $r = 1, 2, \infty$ are covered in [2], we only provide the adaptive algorithms for f_3 and f_4 functions.

- Iterative algorithm for f_3:
The subgradient based adaptive algorithm maximizing $\bar{J}(\boldsymbol{W})$ using the function f_3 can be written as

$$\boxed{\begin{aligned}
&\boldsymbol{W}^{(i+1)} = \boldsymbol{W}^{(i)} \\
&+ \mu^{(i)}\Bigg(\left(\boldsymbol{W}^{(i)}\hat{\boldsymbol{R}}_{\boldsymbol{y}}\boldsymbol{W}^{(i)^T}\right)^{-1}\boldsymbol{W}^{(i)}\hat{\boldsymbol{R}}_{\boldsymbol{y}} - \\
&\frac{3}{2}\sum_{m=1}^{3} g_m \mathbf{e}_m \left(\boldsymbol{y}(l_m^{max(i)}) - \boldsymbol{y}(l_m^{min(i)})\right)^T\Bigg)
\end{aligned}}$$

where $g_m = \dfrac{\hat{\mathcal{R}}(o_1^{(i)}) + \hat{\mathcal{R}}(o_2^{(i)}) + \hat{\mathcal{R}}(o_3^{(i)}) - \hat{\mathcal{R}}(o_m^{(i)})}{\hat{\mathcal{R}}(o_1^{(i)})\hat{\mathcal{R}}(o_2^{(i)}) + \hat{\mathcal{R}}(o_1^{(i)})\hat{\mathcal{R}}(o_3^{(i)}) + \hat{\mathcal{R}}(o_1^{(i)})\hat{\mathcal{R}}(o_3^{(i)})}$ and $l_m^{max(i)}$ ($l_m^{min(i)}$) is the sample index for which the maximum (minimum) value for the m^{th} separator output is achieved at the i^{th} iteration.

- Iterative algorithm for f_4:
The subgradient based adaptive algorithm maximizing $\bar{J}(\boldsymbol{W})$ using the function f_4 can be written as

$$\boxed{\begin{aligned}
&\boldsymbol{W}^{(i+1)} = \boldsymbol{W}^{(i)} \\
&+ \mu^{(i)}\Bigg(\left(\boldsymbol{W}^{(i)}\hat{\boldsymbol{R}}_{\boldsymbol{y}}\boldsymbol{W}^{(i)^T}\right)^{-1}\boldsymbol{W}^{(i)}\hat{\boldsymbol{R}}_{\boldsymbol{y}} - 2p \\
&\sum_{m=1}^{p} \frac{e^{\hat{\mathcal{R}}(o_m^{(i)})}}{log\,(h)\, h}\mathbf{e}_m \left(\boldsymbol{y}(l_m^{max(i)}) - \boldsymbol{y}(l_m^{min(i)})\right)^T\Bigg)
\end{aligned}}$$

where $h = e^{\hat{\mathcal{R}}(o_1^{(i)})} + e^{\hat{\mathcal{R}}(o_2^{(i)})} + \ldots + e^{\hat{\mathcal{R}}(o_p^{(i)})}$.

3. NUMERICAL EXAMPLES AND CONCLUSION

We consider a scenario of 3 sources and 5 mixtures with an i.i.d. Gaussian mixing system. We generate the sources through the Copula-t distribution whose correlation matrix parameter is given by a Toeplitz matrix $\mathbf{R}_s$ whose first row is $\begin{bmatrix} 1 & \rho_s & \ldots & \rho_s^{p-1} \end{bmatrix}$.

Figure 1 shows the output total Signal energy to total Interference energy (over all outputs) Ratio (SIR) obtained for the BCA algorithm examples (corresponding to $f_1, f_{2,1}, f_3, f_4$) for various correlation parameters $\rho_s \in [0, 0.9]$ for the mixture length of $N = 100000$. The same procedure is repeated for FastICA [4], [5] and JADE [6], [7] algorithms, as representative ICA approaches.

In the second example, we generate the sources from exponentially distributed random variables by the inverse CDF method used on the first setup. Figure 2 illustrates the separation performances when the mixture length is $N = 10000$.

Note that the BCA algorithms maintain high separation performance for a wide range of correlation parameters. However, both FastICA and JADE algorithms' performances degrades substantially along with increasing correlation since the independence assumption does not hold. We also point out that for $\rho_s = 0$, the performances of the BCA algorithms

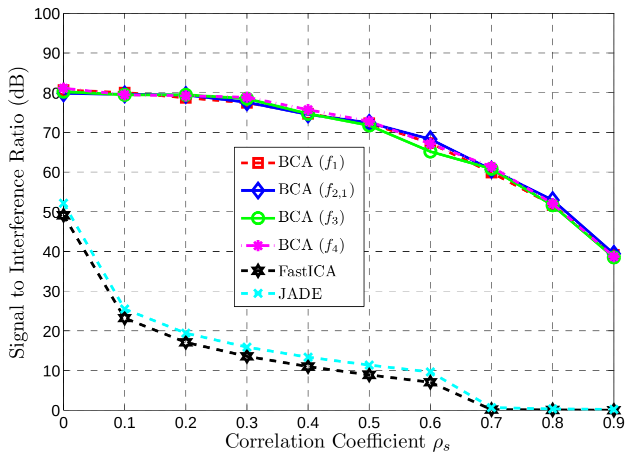

Fig. 1. Result of the proposed BCA algorithms' performances for the mixtures of dependent sources for various correlation parameters when the mixture length is 100000.

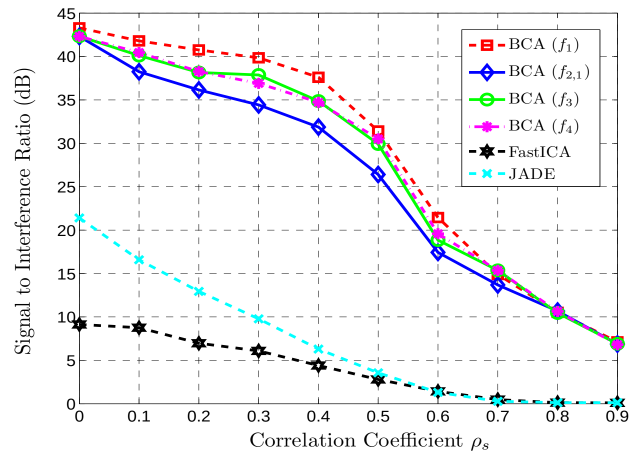

Fig. 2. Result of the proposed BCA algorithms' performances for the mixtures of exponentially distributed dependent sources for various correlation parameters when the mixture length is 10000.

are better than FastICA and JADE even though the independence assumption holds. This is due to the fact that the sample sizes are sufficient for the assumption (**A1**) to hold, whereas they may not be sufficient to reflect the stochastic independence of the sources. We observe that in the second example proposed BCA algorithms have different performances, therefore, the variety of BCA algorithms which can be produced from this analysis might be useful in different scenarios. We finally note that exponential distribution decreases the likeliness of (**A1**) to hold, however, proposed BCA algorithms still have good separation performances and with the longer data records the BCA algorithms become more successful in separating correlated sources.

4. REFERENCES

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

[2] A.T. Erdogan, "A class of bounded component analysis algorithms for the separation of both independent and dependent sources," *Signal Processing, IEEE Transactions on*, vol. 61, no. 22, pp. 5730–5743, Nov 2013.

[3] D.J.H. Garling, *Inequalities: a journey into linear analysis*, Cambridge University Press, 2007.

[4] Aapo Hyvärinen, Juha Karhunen, and Erkki Oja, *Independent Component Analysis*, John Wiley and Sons Inc., 2001.

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

[6] J.-F. Cardoso and A. Souloumiac, "Blind beamforming for non-gaussian signals," *Radar and Signal Processing, IEEE Proceedings F*, vol. 140, no. 6, pp. 362–370, 1993.

[7] Jean-François Cardoso, "High-order contrasts for independent component analysis," *Neural Comput.*, vol. 11, no. 1, pp. 157–192, Jan. 1999.