INTERSPEECH 2017
August 20–24, 2017, Stockholm, Sweden

A Contrast Function and Algorithm for Blind Separation of Audio Signals

Wei Gao, Roberto Togneri, Victor Sreeram

School of Electrical/Electronic and Computer Engineering, The University of Western Australia, Australia

wei.gao@research.uwa.edu.au, roberto.togneri@uwa.edu.au, victor.sreeram@uwa.edu.au

Abstract

This paper presents a contrast function and associated algorithm for blind separation of audio signals. The contrast function is based on second-order statistics to minimize the ratio between the product of the diagonal entries and the determinant of the covariance matrix. The contrast function can be minimized by a batch and adaptive gradient descent method to formulate a blind source separation algorithm. Experimental results on realistic audio signals show that the proposed algorithm yielded comparable separation performance with benchmark algorithms for speech signals, and outperformed benchmark algorithms for music signals.

Index Terms: blind source separation, audio signals, partially correlated, independent component analysis

1. Introduction

Bind Source Separation (BSS) is concerned with recovering unobservable sources from observable mixtures only [1], as the name “blind” suggests. The “blind” feature makes BSS an exciting challenge [2], and is generally compensated by some side information, for instance assumptions about the mixing models and the sources. The weaker the assumptions, the broader the applicability [3], but the harder the BSS problem.

The BSS problem was proposed originally as the “cocktail party problem” on the application of audio signals [2], and attracted much interest for separation of speech signals in the past three decades. A wealth of assumptions and corresponding algorithms have been developed in blind separation of audio signals, in particular speech. Independent Component Analysis (ICA) is one main family of BSS algorithms, which rely on the assumption of statistical independence of the sources. Numerous ICA algorithms of higher order statistics (HOS) and second order statistics (SOS) were developed [4], for example FastICA [5]. ICA algorithms were further developed by incorporating geometric techniques into the implementation of the ICA algorithms. For instance, SOBI [6], STOTD [7] and JADE [8][9] combine Jacobi rotation with second-order, third-order and fourth-order cumulants, respectively.

Another main avenue for blind separation of audio signals is Sparse Component Analysis (SCA) [2]. SCA generally involves sparsifying transformation and clustering algorithms to transform the sources to be as sparse as possible and then detect local dominant bins. Finally, the sources are recovered based on these local dominant bins [10][11]. Moreover, the convex geometric techniques for hyperspectral image unmixing were incorporated into signal processing [12], and the Successive Projection Algorithm (SPA) and the Volume Minimization (VolMin) were formulated by exploiting convex geometry in the covariance domain [13].

At a theoretical level, the audio sources can be considered as statistically independent if the samples are infinite, but in practical scenarios only finite samples of sources are observed. Hence, BSS solutions are estimated from sample-based distributions, which introduce stochastic errors [3]. Speech signals can be considered as nearly uncorrelated, resulting from the fact that natural speech always includes pauses over time and even during voice activity. However, music does not include as many pauses as speech on average, and the harmonic consistency with music may increase the partial correlation, especially on the basis of finite samples. Compared with speech signals, which are nearly W-disjoint orthogonal [10], music usually is less sparse than speech, which hampers the blind separation by SCA algorithms.

In this paper, we consider a standard BSS model without any prior assumptions on the statistical properties of the sources, other than second-order pair-wise correlation. As our simple BSS model does not require additional information on the properties of the sources, it has the potential for more general applications. The remainder of the paper is organized as follows. Section 2 introduces the BSS model and some notation. Section 3 introduces a criterion function, which is extended to a contrast function. Section 4 optimizes the contrast function by a gradient descent method to formulate a BSS algorithm. Section 5 presents the experimental results to evaluate the proposed algorithm. Finally, the conclusions are presented in Section 6.

2. Problem Formulation and Notation

We consider a linear instantaneous BSS model without noise, wherein the observations are determined mixtures of the sources. Sampling at time t, we observe that

$$x(t) = As(t) \tag{1}$$

where $s(t) = [s_1(t), s_2(t), \dots, s_N(t)]^T$ denotes a random vector from N sources, $x(t) = [x_1(t), x_2(t), \dots, x_M(t)]$ denotes a random vector from M observations, and $A \in \mathbb{R}^{M\times N}$ denotes the mixing matrix. Here $t = 1, \dots, L$, and at a theoretical level the number of samples L can be assumed to be positive infinite, while in practical applications L is a finite integer. The BSS model can be represented compactly as

$$X = AS \tag{2}$$

where $X \in \mathbb{R}^{M\times L}$ and $S \in \mathbb{R}^{N\times L}$ denote the mixtures and the sources, respectively. The L samples of the ith source are expressed as S_i, which is the ith row of S. Similarly, X_i and $(\cdot)_i$ denote the ith row of X and the enclosed matrix, respectively. Additionally, $cov(\cdot)$ and $det|\cdot|$ express the covariance and the determinant, respectively.

Due to unavoidable ambiguities of BSS, the goal of BSS is to recover sources $Y = BX = GS \in \mathbb{R}^{N\times L}$ up to a permutation and scaling. Here B represents the separating matrix, and $G = BA \in \mathbb{R}^{N\times N}$ represents the overall mapping. This goal can be posed as finding the separating matrix B such that

http://dx.doi.org/10.21437/Interspeech.2017-189

$G = BA = DP$, where D is a diagonal matrix with non-zero diagonal entries, and P is a permutation matrix.

The audio signals considered in this paper were emitted from independent sources, so they can be assumed pairwise statistically independent at the theoretical level. However, for realistic audio signals that are sample-based, their cross-correlations are usually not zero due to stochastic bias. Considering the properties of realistic audio signals, we assume that

A1: the BSS model is determined, which follows that the mixing matrix A is full-column rank.

A2: Sources are real-valued zero-mean random variables.

A3: Sources are uncorrelated.

We only consider the determined ($M = N$) BSS model in this paper, which can be easily extended to the over-determined ($M > N$) BSS model. Speech is often weak and intermittent [13], so speech signals approximately satisfy **A3**. However, music signals usually do not satisfy **A3**, and are partially correlated. We will show that we can weaken **A3** to partially correlated sources, for music signals.

3. The Proposed Contrast Function

Instead of deriving the contrast function from statistical independence of the sources as the existing ICA algorithms, we consider the overall mapping G, as Erdogan did to formulate a BCA contrast function [14]. Exploiting the assumption of uncorrelated sources, we can formulate a criterion function first based on G as follows:

$$min\{CR(GS)\} = min\left\{\frac{\prod\limits_{i=1}^{N}\Lambda_i\left[cov(GS)\right]}{det|cov(GS)|}\right\} \tag{3}$$

where $\Lambda_i[\cdot]$ returns the (i,i)th diagonal entry of the enclosed square matrix. For the BSS problem, the sources S have been fixed when they are observed. Thus, $CR(GS)$ is a function of the variable G.

Let $\bar{G}$ a perfect solution to distinguish from an imperfect solution $\underline{G}$. A perfect solution can be represented as $\bar{G} = DP$ due to unavoidable ambiguities of BSS, and $\bar{G}$ is full rank by the assumption **A1**.

Firstly, we analyse P in the criterion function. For the denominator of $CR(PS)$ we have $det|cov(PS)| = (det|P|)^2 det|cov(S)| = det|cov(S)|$, given the determinant of a full-rank permutation matrix $det|P| = \pm 1$. For the numerator of $CR(PS)$, we have $\prod\limits_{i=1}^{N}\Lambda_i\left[cov(PS)\right] = \prod\limits_{i=1}^{N} cov(S_i) = \prod\limits_{i=1}^{N}\Lambda_i\left[cov(S)\right]$. Combining the denominator and numerator of $CR(PS)$ together, we can conclude that

$$CR(PS) = CR(S) \tag{4}$$

Secondly, we consider the diagonal matrix D, which is also full-rank. For simplicity and without loss of generality, we can express $D = diag[d_1, \ldots, d_N]$, where $diag[\ ]$ returns a square diagonal matrix with the entries of the enclosed vector on the main diagonal. For the denominator of $CR(DS)$, we have $det|cov(DS)| = (det|D|)^2\ det|cov(S)| = \prod\limits_{i=1}^{N}(d_i)^2\ det|cov(S)|$. For the numerator of $CR(DS)$, we have $\prod\limits_{i=1}^{N}\Lambda_i\left[cov(DS)\right] = \prod\limits_{i=1}^{N} cov(d_i S_i) = \prod\limits_{i=1}^{N}(d_i)^2\ \prod\limits_{i=1}^{N}\Lambda_i\left[cov(S)\right]$. Combining the denominator and the numerator of $CR(DS)$ together, we can conclude that

$$CR(DS) = \frac{\prod\limits_{i=1}^{N}(d_i)^2\prod\limits_{i=1}^{N}\Lambda_i\left[cov(S)\right]}{\prod\limits_{i=1}^{N}(d_i)^2\ det|cov(S)|} = CR(S) \tag{5}$$

Considering that the equality of (4) and (5) rely on the assumption **A1**, we can have

$$CR(\bar{G}S) = CR(DPS) = CR(S) \tag{6}$$

The above analysis demonstrates that the criterion function is preserved for any permutation and scaling. Therefore, for simplicity and without loss of generality, we discuss a simplified G irrespective of the permutation and scaling. We take a simple example of the imperfect solution that only one recovered sources is not separated perfectly, that is

$$\underline{G} = \begin{bmatrix} \cdot & & & & \\ & 1 & & k & \\ & & \cdot & & \\ & & & 1 & \\ & & & & \cdot \end{bmatrix} \tag{7}$$

where $k \neq 0$ at the position (u, v), other off-diagonal entries equal zero, and diagonal entries equal one. Thus, $\underline{G}$ satisfy **A1**. For the denominator of $CR(\underline{G}S)$, we have $det|cov(\underline{G}S)| = det|cov(S)|$, given $det|\underline{G}| = 1$. For the numerator of $CR(\underline{G}S)$, we have

$$\prod_{i=1}^{N}\Lambda_i[cov(\underline{G}S)] = cov(S_u + kS_v)\prod_{i\neq u} cov(S_i) \tag{8}$$

and $cov(S_u + kS_v) = cov(S_u) + 2kcov(S_u, S_v) + k^2 cov(S_v)$, where $cov(S_u, S_v)$ denotes the cross-covariance of row vectors S_u and S_v. The assumption **A2** implies that any source is not a constant, so $cov(S_u) > 0$ and $cov(S_v) > 0$. The assumption **A3** leads to $cov(S_u, S_v) = 0$, which can be approximately satisfied by realistic speech signals. Thus, under **A1**, **A2** and **A3**, we can conclude that $cov(S_u + kS_v) > cov(S_u)$. This can be inserted into (8), and we get that

$$\prod_{i=1}^{N}\Lambda_i[cov(\underline{G}S)] > \prod_{i=1}^{N}\Lambda_i[cov(S)] \tag{9}$$

Combining (9) and the denominator of $CR(\underline{G}S)$ together, we can conclude that

$$CR(\underline{G}S) > CR(S) \tag{10}$$

Based on this inequality for the simplified imperfect solution, it is reasonable to estimate that the inequality holds for a general imperfect solution. Comparing (6) and (10), it is reasonable to estimate that the G obtained by minimizing $CR(GS)$ is a perfect solution, under the assumptions **A1**, **A2** and **A3**.

We further consider partially correlated sources, that is $cov(S_u, S_v) \neq 0$. The inequality of (10) will still hold, if

$$2kcov(S_u, S_v) + k^2 cov(S_v) > 0 \tag{11}$$

By the assumption **A2** and $k \neq 0$, we have $k^2 cov(S_v) > 0$. Thus, if the signs of $cov(S_u, S_v)$ and k are the same, (11) holds. Even though for the worst situation that the signs of $cov(S_u, S_v)$ and k are different, the definitely positive autocorrelation item $k^2cov(S_v)$ contributes to trade off the cross-correlation item $2kcov(S_u, S_v)$, when the sources are partially correlated.

Additionally, robustness is increased by processing series partitioned observations rather than the observations as a matrix, as [6] analysed. We can partition S under an l-length time window and let $S[w] = [s(wl+1), \dots, s(wl+l)] \in \mathbb{R}^{N \times l}$ denote the wth partitioned time window. Thus, we can incorporate the time windows into the criterion function (3) to propose a contrast function as

$$CF(GS[w]) = \prod_{w \in W} \{CR(GS[w])\} \qquad (12)$$

where W denotes the set of time windows, and the time window can be overlapped or non-overlapped. Observe that if the window length l is small, in some time windows $det|cov(S[w])| \approx 0$. It follows that the denominator of the contrast function is nearly zero within these time windows, and these time windows should be removed from the calculation of the contrast function. Because S is unobservable, we filter off the time window when $det|cov(X[w])| \approx 0$, which is equivalent to $det|cov(S[w])| \approx 0$, given $det|A| \neq 0$ by the assumption **A1**. Hence, we refer to W as only containing time windows such that $det|cov(X[w])| \neq 0$. The choice of l is important, which we will discuss in Section 5.1.

We compare our proposed contrast function with the existing ICA algorithms. On the basis of statistical independence, SOBI zeros off-diagonal entries of several covariance matrices by Jacobian rotation. However, for some signals their cross-covariances are not zero due to stochastic bias or harmonic consistency, so the separation performance of SOBI is likely to suffer for these sources. In the same way, the cross higher-order statistics of realistic sources are not zero, so zeroing these off-diagnal entries of higher-order statistical matrices [15], for example STOTD [7] and JADE [8][9], leads to separation bias. The sample-based bias usually lead to the sources are partially correlated. Given the sources are at most partially correlated, our contrast function exploits the autocorrelation of a source to trade off the corresponding cross-correlation. Hence, the trade-off effect help us to weaken the assumption **A3** to partially correlated sources, and our proposed function have the potential to separate partially correlated sources.

4. The Proposed Algorithm

In the previous section, we derived the contrast function (12) of G, but for the BSS problem the mixing matrix A is unavailable and only the observations X are available. Firstly, we recast the contrast function to relying on X and updating the separating matrix B as $CF(BX[w]) = CF(BAS[w]) = CF(GS[w])$. This section presents an associated algorithm to minimize the modified contrast function. Since the contrast function is first-order differentiable, we exploit a gradient descent method to minimize the contrast function.

Furthermore, for computational simplicity of the first-order derivative, the contrast function can be modified to $log(\sqrt{CF(BX[w])})$, due to the monotonicity of the power root function and the natural logarithm function. Thus, a perfect solution will be obtained by

$$\bar{B} = \arg\min_{\mathbf{B}}\{log(\sqrt{CF(BX[w])})\} \qquad (13)$$

Numerous BSS algorithms use whitening as a preprocessing step to denoise or transform from an over-determined case to a determined case [2]. Whitening exploits eigenvalue decomposition of the covariance matrix $cov(X)$ or singular value decomposition of the observations X to transform X by a whitening matrix Q into an uncorrelated matrix $Z = QX$ such that $cov(QX) = I$, where I is an identity matrix. For more details of whitening a matrix see [2][5]. Our algorithm applies whitening to obtain $Z = QX$ first, and we can thus set the initial value $B^{(1)} = I$.

$$\begin{aligned} \textbf{Let } C_Z[w] &= cov(Z[w]), \\ DI(w) &= \sum_{i=1}^{N} log(\Lambda_i \left[BC_Z[w]B^T\right]), \\ DE(w) &= log(det|BC_Z[w]B^T|). \end{aligned}$$

$$\therefore log(\sqrt{CF(BZ[w])}) = \sum_{w \in W} 0.5\{DI(w) - DE(w)\} \qquad (14)$$

Within a time window w, we have

$$\frac{\partial\{0.5DI(w)\}}{\partial B} = \begin{bmatrix} \frac{B_1C_Z[w]}{B_1C_Z[w]B_1^T} \\ \cdot \\ \frac{B_NC_Z[w]}{B_NC_Z[w]B_N^T} \end{bmatrix} \qquad (15)$$

$$\frac{\partial\{0.5DE(w)\}}{\partial B} = \left[BC_Z[w]B^T\right]^{-1} BC_Z[w] \qquad (16)$$

where B_i denotes the ith row of B. Combining (15) and (16) together, we get the gradient of (14) within a time window w as

$$Gra[w] = \begin{bmatrix} \frac{B_1C_Z[w]}{B_1C_Z[w]B_1^T} \\ \cdot \\ \frac{B_NC_Z[w]}{B_NC_Z[w]B_N^T} \end{bmatrix} - \left[BC_Z[w]B^T\right]^{-1} BC_Z[w] \qquad (17)$$

Therefore, by the gradient descent method, we can update B by a batch of gradients as

$$B^{(U+1)} = B^{(U)} - \eta^{(U)} \sum_{w \in W} Gra[w] \qquad (18)$$

where $\eta^{(U)}$ denotes the Uth step size. The step size affects convergence and computational burden. Typically an adaptive step size is better than a fixed step size, and one widely used method for calculating adaptive step sizes is linear search, which usually increases computational cost. To balance computational burden and convergence, we set the adaptive step size as

$$\eta^{(U)} = \frac{log(\sqrt{CF(B^{(U)}Z[w])}\,)}{norm(\sum_{w \in W} Gra[w])} \qquad (19)$$

where $norm(\cdot)$ denotes the Frobenius norm of the enclosed matrix. Additionally, we optimize the contrast function by summing up gradients across several time windows, as a batch gradient descent method. The batch gradient can mitigate the zig-zag phenomenon of the gradient descent method to speed up convergence. Furthermore, the denominator $det|cov(GS)| = (det|G|)^2 det|S| = (det|D|)^2 det|S|$ cancels the scaling $(det|D|)^2$ in the numerator of the contrast function. Hence, we do not need to normalize $B^{(U)}$ in each updating iteration.

Finally, the separating matrix can be obtained from $B = B^{(U)}Q$ due to the whitening [2]. The recovered sources are then $Y = BX/norm(B)$, where a scaling $norm(B)$ is applied. The pseudo code of the proposed algorithm is presented in Table 1.

Table1: Pseudo Code of the Proposed Algorithm

```
Step1: Whiten observations $Z = QX$, and initialize $B^{(1)} = I$
Step2: Calculate $C_Z[w] = cov(Z[w]), w \in W$
    Filter off $\forall w$ such that $det|cov(Z[w])| \approx 0$
Step3: Update separating matrix $B^{(U)}$ until convergent
    Step3.1: Calculate gradient $\sum_{w \in W} Gra[w]$ (17)
    Step3.2: Calculate the step size $\eta$ in (19)
    Step3.3: Update $B^{(U+1)}$ in (18)
Step4: Results $B = B^{(U)}Q$, and then $Y = BX/norm(B)$.
```

5. Experimental Results

5.1. Experimental Settings

In order to evaluate our proposed algorithm, we present experimental results for both realistic speech and music signals. Our speech dataset contains 50 speech segments from the TIMIT database of length 6 s and frequency 16 KHz. Our music dataset contains 70 piano [1] or violin [2] solos of length 6 s and frequency 44.1 KHz. In each simulation, 200 independent runs were performed and averaged. For each independent run, the sources were randomly selected from our datasets, and the entries of A were generated randomly from a zero-mean unit-variance normal distribution.

We benchmarked the proposed algorithm against the ICA algorithms: FastICA [5], SOBI [6] and JADE [8], and two sparsity-based algorithms: SPA and VolMin [13]. The BSS performance index PI [16] has been used as an evaluation measure:

$$PI(G) = \frac{1}{2N(N-1)} \left\{ \sum_{i=1}^{N} \left(\sum_{j=1}^{N} \frac{|G(i,j)|^2}{max_k |G(i,k)|^2} - 1 \right) + \sum_{j=1}^{N} \left(\sum_{i=1}^{N} \frac{|G(i,j)|^2}{max_k |G(k,j)|^2} - 1 \right) \right\}$$

A smaller PI implies a better separation performance. One issue we have not addressed is the window length l. If the window length is small, the sources are more likely to satisfy the assumption **A3**. However, the determinant of a covariance matrix within a small time window is more likely to be zero. These nearly zero-determinant time windows will be filtered off for the calculation of the contrast function. Additionally, the small window length generates more time windows, which usually increase computational cost. In contrast, if the window length is large, the total computational cost may be reduced, but the sources are less likely to satisfy the assumptions **A3**. Our experiments show that the best results were obtained with window lengths between the interval $[500, 1000]$. To decrease computational burden, this paper exploits non-overlapped time windows.

5.2. Experimental Results

Fig. 1 depicts the PI of speech sources recovered by the proposed algorithm with different window lengths of $w = 500, 800, 1000$ and the number of sources $N = 3, \dots, 10$. We can see that for 8 sources and less, the proposed algorithm yielded better separation performance than the benchmarks. In general, the separation performance of all these algorithms was under $-30dB$, which can be considered satisfactory in practical applications. In fact, blind separation of speech signals has been well-developed for the determined BSS model, because speech signals can approximately satisfy the assumption **A3** and W-disjoint orthogonal.

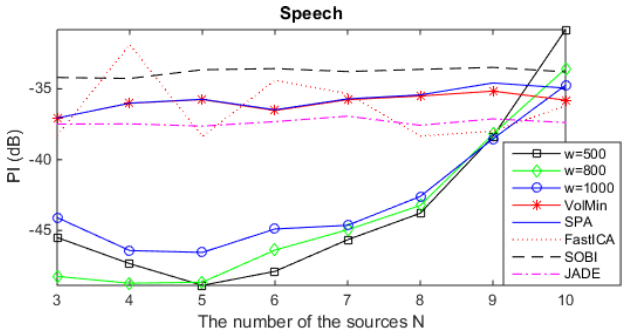

Figure 1: *Mean PI for speech by the number of sources N. The results of the proposed algorithm are labelled as "w=500", "w=800" and "w=1000" corresponding to the window lengths.*

Our focus here is on music signals. Fig. 2 shows the PI of music signals with different window lengths of $w = 500, 800, 1000$ and the number of sources $N = 3, \dots, 10$ as before. As expected the proposed algorithm outperformed all the benchmark algorithms, although the overall separation performance was worse than that obtained with speech. This is because music signals are more likely to be partially correlated rather than uncorrelated as is the case with speech. The separation performance of our proposed algorithm was under $-21dB$, which in practical applications can be considered promising. These results support that our contrast function is robust to the partial correlation of the sources.

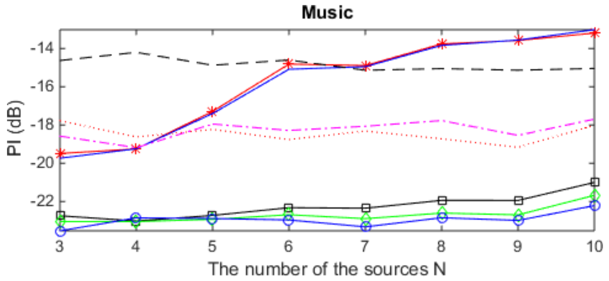

Figure 2: *Mean PI for music by the number of sources N. The legend is the same as the legend of Fig.1.*

6. Conclusions

In this paper, we analyse the overall mapping G to introduce a contrast function, which is not based on measuring independence as in the conventional ICA algorithms. Exploiting the autocorrelation of the sources to trade off the cross-correlation, the contrast function is robust to the partial correlation of the sources. A BSS algorithm is presented by a batch gradient descent method to optimize the contrast function. Experimental results show that the proposed algorithm outperformed benchmarks for music signals.

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

[2] http://www.tasminlittle.org.uk

7. References

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

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

[3] J. F. Cardoso, “Blind signal separation: Statistical principles,” *IEEE Proc.*, vol. 86, no. 10, pp. 2009–2025, 1998.

[4] D. T. Pham, “Blind separation of instantaneous mixture of sources via an independent component analysis,” *IEEE Trans. Signal Process.*, vol. 44, no. 11, pp. 2768–2779, 1996.

[5] A. Hyvarinen, “Fast and robust fixed-point algorithms for independent component analysis,” *IEEE Trans. Neural Netw.*, vol. 10, no. 3, pp. 626–634, 1999.

[6] A. Belouchrani, K. AbedMeraim, J. F. Cardoso, and E. Moulines, “A blind source separation technique using second-order statistics,” *IEEE Trans. Signal Process.*, vol. 45, no. 2, pp. 434–444, 1997.

[7] L. De Lathauwer, B. De Moor, and J. Vandewalle, “Independent component analysis and (simultaneous) third-order tensor diagonalization,” *IEEE Trans. Signal Process.*, vol. 49, no. 10, pp. 2262–2271, 2001.

[8] J. F. Cardoso and A. Souloumiac, “Blind beamforming for non-gaussian signals,” *IEE Proc. F (Radar and Signal Process.)*, vol. 140, no. 6, pp. 362–370, 1993.

[9] J.-F. Cardoso, “On the performance of orthogonal source separation algorithms,” in *Proc. EUSIPCO*, Edinburgh, Sep. 1994, pp. 776–779.

[10] O. Yilmaz and S. Rickard, “Blind separation of speech mixtures via time-frequency masking,” *IEEE Trans. Signal Process.*, vol. 52, no. 7, pp. 1830–1847, 2004.

[11] V. G. Reju, S. N. Koh, and I. Y. Soon, “Underdetermined convolutive blind source separation via time-frequency masking,” *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 18, no. 1, pp. 101–116, 2010.

[12] W. K. Ma, J. M. Bioucas-Dias, T. H. Chan, N. Gillis, P. Gader, A. J. Plaza, A. Ambikapathi, and C. Y. Chi, “A signal processing perspective on hyperspectral unmixing,” *IEEE Signal Process. Mag.*, vol. 31, no. 1, pp. 67–81, 2014.

[13] X. Fu, W.-K. Ma, K. Huang, and N. D. Sidiropoulos, “Blind separation of quasi-stationary sources: Exploiting convex geometry in covariance domain,” *IEEE Trans. Signal Process.*, vol. 63, no. 9, pp. 2306–2320, 2015.

[14] A. T. Erdogan, “A class of bounded component analysis algorithms for the separation of both independent and dependent sources,” *IEEE Trans. Signal Process.*, vol. 61, no. 22, pp. 5730–5743, 2013.

[15] D. T. Pham and J. F. Cardoso, “Blind separation of instantaneous mixtures of nonstationary sources,” *IEEE Trans. Signal Process.*, vol. 49, no. 9, pp. 1837–1848, 2001.

[16] A. Boudjellal, A. Mesloub, K. Abed-Meraim, and A. Belouchrani, “Separation of dependent autoregressive sources using joint matrix diagonalization,” *IEEE Signal Proc. Lett.*, vol. 22, no. 8, pp. 1180–1183, 2015.