# BLIND SOURCE SEPARATION VIA A WEAK EXCLUSION PRINCIPLE

*Zhang Zihan, Thierry Blu*

The Chinese University of Hong Kong

## ABSTRACT

In this paper, we propose a generalized Blind Source Separation (BSS) method using a novel assumption which we call "weak exclusion" principle. We first give the mathematical definition of the exclusion criterion and propose an iterative algorithm to minimize it. We then test WEP in simulated and real datasets, compared with other four methods. The experiments on synthetic and real datasets demonstrate that WEP outperforms the other methods, both in terms of accuracy and in terms of speed.

***Index Terms*—** Blind Source Separation, Weak Exclusion Principle, Hyperspectral Images Unmixing

## 1. INTRODUCTION

Blind source separation (BSS) is a kind of data-driven problem, which separates underlying source signals and corresponding mixing relationship from the observed signals [1]. In the BSS problem, minimal prior knowledge and flexible assumptions are utilized to complete the separation task [2]. It emerged from the classical cocktail-party problem in the field of audio signal processing [3]. BSS has a wide range of applications in many fields, such as medical imaging, remote sensing, finance, acoustics, radio communication [4].

In the past few years, the field of blind source separation has mainly focused on the natural dimensional or high-range dimensional characteristics of data, combining diversified prior conditions with different decomposition models to complete the separation task. There are various avenues for research. Independent component analysis (ICA) is the first efficient BSS method, which uses non-Gaussianity and sample dependence in the way of diversity with according mutual information rate as the cost. In terms of methods utilizing statistic properties, the two main directions are maximum likelihood estimation [5] and joint diagonalization [6]. Considering the natural characteristics of the observed signals, such as nonnegativity, nonnegative matrix factorization (NMF) and tensors have become active fields. Various source constraints are applied to the standard or modified NMF and tensor frameworks to build temporal models, such as FastMNMF [7] and KL-NMF [8]. Another widely developed technical route is to decompose the mixed signals according to the sparsity and low rank of the matrix. This type of method uses a series of assumptions to provide the prior knowledge required by the model and performs sparse representation or low-rank representation on the matrix, such as LR-NTF[9]. The deep learning methods that have emerged in recent years also have been applied in BSS, such as Deep S3PR [10]. Due to the data limitations of the BSS task itself, on the one hand, it is difficult to label datasets and calculate fast; on the other hand, it is challenging to transfer the deep model between different mixed datasets.

In this paper, the concept of "exclusion" is meant to describe the ideal BSS situation where only one kind of source exists at each time instant or pixel location. It is easy to extend it to the general situation where one kind of source is significantly larger than the others in each time signal, which is called the Weak Exclusion Principle (WEP). It should be pointed out that exclusion is an underlying property that is implicitly present in the BSS field. A good example is the sparsity assumption. This type of method seeks the most sparse solution, that is, the number of zeros in the source matrix is as large as possible. The optimal norm used in the ideal state is considered to be the $\ell_0$ norm. Because of the NP-hard problem in solving $\ell_0$ norm, $\ell_1$ norm is often used instead for sparse representation [11]. The effect of the maximum number of zeros pursued by sparsity might result in the absence of source distribution in some time signals. But the complete exclusion property can ensure sparseness without causing loss of source data.

In this paper, we propose a theoretical reformulation of the empirical algorithm that we had earlier developed for EEG only[12]. The theory now includes a quantitative definition of the exclusion, the minimization of which results in our algorithm. Moreover, we show how to use this algorithm for images, even for those that do not satisfy the exclusion principle. As far as we know, this is the first application of this criterion to hyperspectral images.

## 2. PROPOSED MATHEMATICAL MODEL

### 2.1. Blind Source Separation

In a blind source separation (BSS) setting, a $K \times L$ matrix $\mathbf{X}$ with $K$ time samples and $L$ signals can be separated as a linear combination of $K \times M$ source matrix $\mathbf{S}$ with $M$ sources using weights from a $M \times L$ mixing matrix $\mathbf{A}$. This model is

described mathematically as

$$\mathbf{X} = \mathbf{SA} \tag{1}$$

where $\mathbf{X} = [x_{k,l}]_{1\leq k\leq K, 1\leq l\leq L}$, $\mathbf{S} = [s_{k,m}]_{1\leq k\leq K, 1\leq m\leq M}$, and $\mathbf{A} = [a_{m,l}]_{1\leq m\leq M, 1\leq l\leq L}$. The goal of BSS is to retrieve $\mathbf{S}$ and $\mathbf{A}$ from the knowledge of $\mathbf{X}$. Without further hypotheses, the solution of BSS is not unique, which has been explained clearly in[1, 13].

### 2.2. Exclusion Principle

We say that a source matrix $\mathbf{S}$ is exclusive iff, for any row index $k$, only one column of $\mathbf{S}$ is non-zero. Equivalently, for any diagonal matrix $\mathbf{W}$, then $\mathbf{S}^{\mathrm{H}}\mathbf{WS}$ is a diagonal matrix as well. The "exclusion" of a $K \times M$ matrix $\mathbf{S}$ is defined as a positive scalar number according to

$$\mathscr{E}\{\mathbf{S}\} \overset{\text{def}}{=} 1 - \frac{1}{M}\sum_{k=1}^{K} \max_{m=1,\ldots,M} |\tilde{s}_{k,m}|^2 \tag{2}$$

where the tilde notation stands for the normalization that transforms $\mathbf{S}$ into a matrix of same dimension defined by $\tilde{\mathbf{S}} = \mathbf{S}\,\mathrm{diag}\left(\mathbf{S}^{\mathbf{H}}\mathbf{S}\right)^{-1/2}$; or, equivalently

$$\tilde{s}_{k,m} \overset{\text{def}}{=} \frac{s_{k,m}}{\sqrt{\sum_{k'=1}^{K} |s_{k',m}|^2}}. \tag{3}$$

Through (2), it can be obtained that $0 \leq \mathscr{E}\{\mathbf{S}\} \leq 1 - 1/M$ and $\mathbf{S}$ is exclusive iff $\mathscr{E}\{\mathbf{S}\} = 0$, which suggests that it is possible to define the exclusion of sources using the quantity $\mathscr{E}\{\mathbf{S}\}$. In that case, weakly exclusive sources are sources for which $\mathscr{E}\{\mathbf{S}\}$ is small. To describe the exclusion directly by source matrix, consider the diagonal $K \times K$ matrices $\{\mathbf{W}_m\}_{1\leq m\leq M}$ whose diagonal elements are given by

$$w_m[k,k] = \begin{cases} 1, & \text{if } |\tilde{s}_{k,m}| \geq |\tilde{s}_{k,m'}|, \forall m' \in \{1,\ldots,M\}; \\ 0, & \text{otherwise .} \end{cases}$$

Then, denoting by $\{\mathrm{e}_m\}_{m=1,\ldots,M}$ the $M \times 1$ canonical basis of vectores of $\mathbb{C}^M$ ($m^{th}$ coordinate equal to 1, and all others equal to 0), we have

$$\mathscr{E}\{\mathbf{S}\} = 1 - \frac{1}{M}\sum_{m=1}^{M} \left\|\mathbf{W}_m \tilde{\mathbf{S}} \mathrm{e}_m\right\|^2 \tag{4}$$

### 2.3. Separation of Exclusive Sources

Assume the matrix $\mathbf{S}$ in Eq.2 is exclusive and choose two different diagonal matrices $\mathbf{W}_1$ and $\mathbf{W}_1$, we have that $\mathbf{S}^{\mathrm{H}}\mathbf{W}_i\mathbf{S} = \mathbf{D}_i (i = 1, 2)$ are diagonal matrices. Then, this means that

$$\mathbf{X}^{\mathrm{H}}\mathbf{W}_1\mathbf{X} = \mathbf{A}^{\mathrm{H}}\mathbf{D}_1\mathbf{A} \quad \text{and} \quad \mathbf{X}^{\mathrm{H}}\mathbf{W}_2\mathbf{X} = \mathbf{A}^{\mathrm{H}}\mathbf{D}_2\mathbf{A} \tag{5}$$

Let us assume that $\mathbf{A}$ is invertible and $\mathbf{D}_1\mathbf{D}_2^{-1} = \mathrm{diag}\,(\lambda_1, \lambda_2, \ldots, \lambda_M)$ has distinct diagonal values ($\lambda_i \neq \lambda_j$, if $i \neq j$). It is observed that $\mathbf{X}^{\mathrm{H}}\mathbf{W}_1\mathbf{X}\mathbf{A}^{-1}\mathbf{D}_2 = \mathbf{X}^{\mathbf{H}}\mathbf{W}_2\mathbf{X}\mathbf{A}^{-1}\mathbf{D}_1$ or, better,

$$\mathbf{X}^{\mathrm{H}}\mathbf{W}_1\mathbf{X}\underbrace{\mathbf{V}}_{\mathbf{A}^{-1}} = \mathbf{X}^{\mathrm{H}}\mathbf{W}_2\mathbf{X}\mathbf{V}\underbrace{\mathbf{D}}_{\mathbf{D}_1\mathbf{D}_2^{-1}}. \tag{6}$$

This shows that $\mathbf{V}$ and $\mathbf{D}$ are the generalized eigenvectors and eigenvalues of the matrices $\mathbf{X}^{\mathrm{H}}\mathbf{W}_1\mathbf{X}$ and $\mathbf{X}^{\mathrm{H}}\mathbf{W}_2\mathbf{X}$, which suggests a simple algebraic solution to the BSS problem. The fact that the generalized eigenvalue decomposition is unique up to scaling ensures the unicity of the matrix $\mathbf{A}^{-1}$.

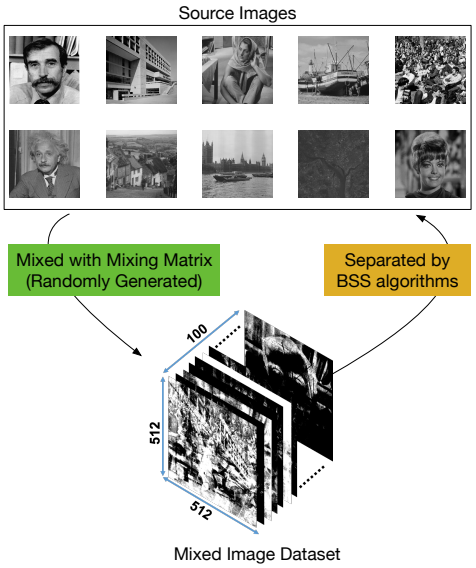


**Fig. 1**. Generation progress of the simulated image dataset.

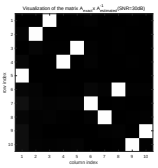

**Fig. 2**. Visualization of the matrix $\mathbf{A}_{exact} \times \mathbf{A}^{-1}_{estimated}$ for SNR = 30 $dB$.

### 2.4. Separation of Weakly Exclusive Sources

Looking for the most exclusive solution of a BSS problem can be formulated through the minimization:

$$\min_{\mathbf{B}} \mathscr{E}\{\mathbf{XB}\}. \tag{7}$$

where $\mathbf{B} = \mathbf{A}^{-1}$, which in turn provides that $\mathbf{S} = \mathbf{XB}$.

***WEP Algorithm***-Solving the non-convex optimization problem Eq.(7) can be done by iteratively:

1. minimizing Eq.(4) under the normalization constraint that $\mathrm{XB} = \widetilde{\mathrm{XB}}$, assuming that the matrices $\mathbf{W}_m$ in Eq.(4) are known;

2. updating $\mathbf{W}_m$. The advantage of knowing matrices $\mathbf{W}_m$ is that the exclusion criterion becomes quadratic in $\mathbf{B}$.

**Table 1**. Numerical Results and Running Time(seconds) on Simulated Separation Dataset. Best Results are Illustrated in Bold.

| SNR | Metric | LR-NTF | GLMM | FCLSU-VCA | NMF-QMV | WEP |
|---|---|---|---|---|---|---|
| 15dB | PI | 0.4722 | 0.3597 | 0.3427 | 0.3566 | **0.0573** |
| | RMSE_A | 0.1805 | 0.1943 | 0.1792 | 0.1776 | **0.1010** |
| | RMSE_S | 0.2800 | 0.2795 | 0.2707 | 0.2156 | **0.1334** |
| | RMSE_I | 0.1896 | 0.1346 | 0.1576 | 0.1687 | **0.1287** |
| | Time | 1025.98 | 960.66 | 26.51 | 883.84 | **4.21** |
| 20dB | PI | 0.4658 | 0.3597 | 0.3427 | 0.2889 | **0.0226** |
| | RMSE_A | 0.1832 | 0.1943 | 0.1792 | 0.1757 | **0.0819** |
| | RMSE_S | 0.2785 | 0.2795 | 0.2707 | 0.1952 | **0.1264** |
| | RMSE_I | 0.1921 | 0.1346 | 0.1576 | 0.1271 | **0.0848** |
| | Time | 1008.65 | 999.51 | 30.33 | 886.98 | **4.11** |
| 30dB | PI | 0.4089 | 0.4209 | 0.3147 | 0.2369 | **0.0177** |
| | RMSE_A | 0.1795 | 0.1998 | 0.1794 | 0.1752 | **0.0792** |
| | RMSE_S | 0.2295 | 0.2791 | 0.2697 | 0.1917 | **0.1343** |
| | RMSE_I | 0.1617 | 0.0967 | 0.1290 | 0.0749 | **0.0438** |
| | Time | 1040.81 | 977.78 | 23.32 | 869.01 | **4.49** |
| 40dB | PI | 0.5123 | 0.3254 | 0.4542 | 0.3714 | **0.0165** |
| | RMSE_A | 0.3030 | 0.2002 | 0.1795 | 0.1757 | **0.0798** |
| | RMSE_S | 0.2434 | 0.2792 | 0.2697 | 0.2007 | **0.1362** |
| | RMSE_I | 0.1690 | 0.1055 | 0.1247 | 0.0540 | **0.0399** |
| | Time | 1041.51 | 1043.22 | 29.08 | 2088.70 | **4.37** |

The optimal solution of this problem is given by

$$\mathbf{X^H W}_m\mathbf{XBe}_m = \lambda_m \mathbf{X^H XBe}_m, \quad m = 1, \ldots, M. \quad (8)$$

where $\lambda_m$ are positive real numbers which should be chosen to maximize $\sum_{m=1}^{M} \lambda_m$ for the reason that this is a set of decoupled generalized eigendecomposition problems. Perform the economy size SVD of $\mathbf{X}$ under the form of $\mathbf{U}\Delta\mathbf{V}^{\mathrm{H}}$($\mathbf{U}$ and $\mathbf{V}$ unitary, $\mathbf{\Delta}$ diagonal ) and introduce the vector variable $\mathbf{Y}_m = \mathbf{\Delta V}^{\mathrm{H}}\mathbf{Be}_m$.

In this way, the generalized eigendecomposition problem simplifies to the eigendecomposition of the matrix $\mathbf{U}^{\mathrm{H}}\mathbf{W}_m\mathbf{U}$, where $\mathrm{Y}_m$ is the eigenvector. So it can be efficiently obtained by performing the SVD of $\mathbf{W}_m\mathbf{U}$. This provides $\mathbf{Y}_m$, which in turn provides $\mathbf{Be}_m = \mathbf{V\Delta}^{-1}\mathbf{Y}_m$. Once $\mathbf{B}$ has been obtained, $\mathbf{S} = \mathbf{XB}$ is estimated and the diagonal matrices $\mathbf{W}_m$ are updated according to their definition in Eq.(4). The algorithm stops iterating when the matrices $\mathbf{W}_m$ do not change anymore.

## 3. EXPERIMENT WITH SIMULATED DATA

A $512 \times 512 \times 100-$size dataset is made by randomly mixing ten $512 \times 512-$size images, which are chosen arbitrarily from the Internet and between which there is no spatial or temporal connection. The mixing relationship is also a randomly formed matrix. The purpose of separation is to decompose these ten pictures from the dataset. The generation progress of this simulated dataset is shown in Figure 1. Four useful algorithms are utilized for comparison: LR-NTF[9], GLMM[14], FCLSU-VCA[15], NMF-QMV[16]. $\mathrm{LR}-\mathrm{NTF}$ and $\mathrm{NMF}-\mathrm{QMV}$ are newly$-$published algorithms. GLMM and $\mathrm{FCLSU}-\mathrm{VCA}$ are classical blind source separation methods with good efficiency.

**Table 2**. Numerical Results and Running Time(seconds) on Hyperspectral Urban Dataset. Best Results are Illustrated in Bold.

| Metric | LR-NTF | GLMM | FCLSU-VCA | NMF-QMV | WEP |
|---|---|---|---|---|---|
| PI | 0.5099 | 0.4237 | 0.4280 | 0.4442 | **0.1149** |
| RMSE_S | 0.0231 | 0.0469 | 0.0326 | 0.0272 | **0.0185** |
| RMSE_A | 0.1067 | 0.0855 | 0.0917 | 0.0845 | **0.0306** |
| RMSE_I | 0.1749 | 0.1409 | 0.1046 | 0.0539 | **0.0454** |
| $\mathscr{E}\{\mathbf{S}_{est}\}$ | 0.4711 | 0.2390 | 0.1646 | 0.4372 | **0.1299** |
| Time | 238.17 | 165.32 | 6.34 | 413.54 | **1.47** |

Three kinds of metrics are used in experiments to measure the separation results of WEP and comparison algorithms: Performance Index(PI), Root-Mean-Square-Error(RMSE) and Running Time. PI is used to evaluate how close the retrieved mixing matrix is to the ground-truth $\mathbf{A}_{\text{exact}}$. A common condition is that the ordering of the sources cannot be retrieved only through the BSS progress. So ideally, what is expected is that the matrix $\mathbf{P} = \mathbf{A}_{exact} \times \mathbf{A}^{-1}_{estimated}$ is diagonal after a suitable permutation of columns. This leads to the definition of PI [17]. When $\mathbf{A}_{estimated}$ is closer to $\mathbf{A}_{\text{exact}}$, PI is smaller.

$$\mathrm{PI} = \frac{1}{M(M-1)} \left( \sum_{m,n=1}^{M} \frac{|p_{m,n}|^2}{\max\limits_{m'=1,\ldots,M} |p_{m,m'}|^2} - M \right) \quad (9)$$

RMSE is utilized to measure the difference between the groundtruth $\mathbf{Q}$ and estimated results $\widehat{\mathbf{Q}}$ as follows[18]:

$$\mathrm{RMSE} = \sqrt{\frac{1}{J \times K} \|\mathbf{Q} - \widehat{\mathbf{Q}}\|_F^2} \quad (10)$$

where $J$ and $K$ are the number of columns and rows of the measured matrixes. Source matrix (RMSE_S), the mixing matrix (RMSE_A), and the mixed matrix (RMSE_I) are used respectively. Four different values of Signal-Noise-Ratio (SNR) is set as: $15dB$, $20dB$, $30dB$ and $40dB$. Given that the exclusion of the source matrix in groundtruth is 0.7637, it should be noted that the simulated dataset is not naturally exclusive, so we need to apply a Laplacian of Gaussian (LoG) filter to the mixture in WEP.

Table 1 illustrates the numerical results of WEP and four comparison algorithms on the simulated dataset. It can be

observed clearly that the metric values of WEP on PI and RMSE are significantly smaller than comparison methods. Figure 2 shows the visualization of the matrix $\mathbf{P}$ in $SNR = 30dB$. The darker the color of the square, the closer the value is to 0, and the brighter the color, the closer the value is to 1. Observing Table 1 and Figure 2, it can be seen that WEP is less affected by different SNR values compared with other algorithms, which shows that WEP has good robustness. Another thing shown in the Table 1 is that WEP is very fast on the simulated image dataset. It benefits from calculating the SVD of a tall matrix, which is fast.

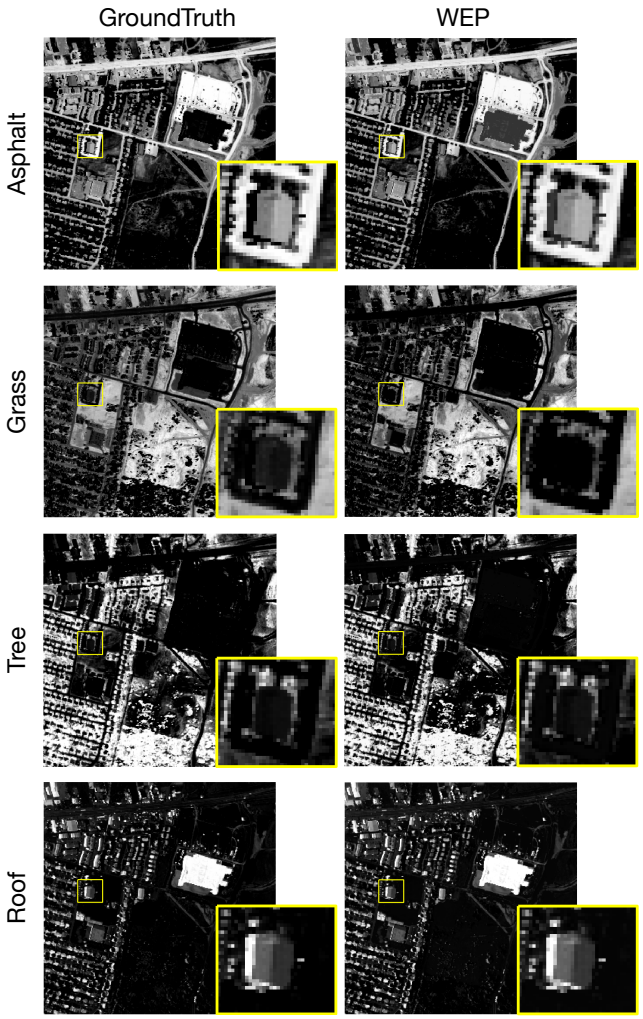


**Fig. 3**. Experimental results on Urban dataset. Parts of the source images are blocked with yellow frames and enlarged to show the detailed separation performance.

## 4. EXPERIMENT WITH REAL DATA

Blind source separation has many practical applications, one of which is hyperspectral image unmixing. In order to test the performance of the proposed algorithm in the real dataset, Urban, a hyperspectral dataset with groundtruth provided in [19], is utilized for separation. The size of Urban is $307 \times 307 \times 162$ with 4 kinds of sources or endmembers named in unmixing. Same comparison algorithms and metrics as the simulated dataset are employed in the Urban dataset. Given that the exclusion value of source matrices in the Urban is 0.0653, it is obvious that Urban dataset is exclusive. So the metric $\mathscr{E}\{\mathbf{S}_{est}\}$ can be used here to measure the exclusion of estimated results.

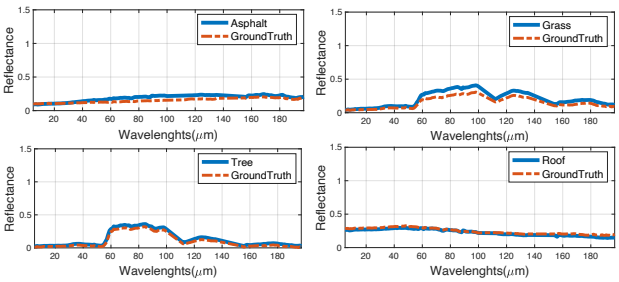


**Fig. 4**. The spectral wavelengths of groundtruth (dotted orange lines) and mixing matrices (solid blue lines) estimated by WEP.

Table 2, Figure 3 and Figure 4 show the quantitative and visual results obtained by WEP and comparison algorithms on Urban dataset. It should be pointed out that the results of comparison algorithms on the real dataset are better than the results on the simulated data. There are two main reasons. On the one hand, comparison algorithms contain many specific constraints for the task of hyperspectral unmixing. On the other hand, source matrices of the real dataset are spatially related, which is randomly mixed in the simulated dataset. The separation is less difficult in the Urban dataset compared with the above simulated dataset.

It is important to note that WEP, as a generalized blind source separation theory, does not add any specific constraints for hyperspectral datasets. However, it can be observed on Table 2 that PI and RMSE of WEP are the lowest among all the results. This shows that WEP can be used in a wide range of practical applications without being restricted by data types. It can be noticed that the smaller the exclusion value of the separation source results, the better the algorithm works. That is to say, it verifies the assumption that the retrieval and utilization of exclusion is the underlying unified theory of BSS algorithms.

## 5. CONCLUSION

This paper proposes a BSS method using exclusion. The main advantages are listed as follows: no additional prior knowledge is required; it is a general theory for different applications; the calculation is very fast. Experiments prove that it gives excellent results on different data sets.

## 6. REFERENCES


[1] Tülay Adalı, Christian Jutten, Arie Yeredor, Andrzej Cichocki, and Eric Moreau, "Source separation and applications," *IEEE Signal Processing Magazine*, 2014.

[2] Te-Won Lee, Michael S Lewicki, Mark Girolami, and Terrence J Sejnowski, "Blind source separation of more sources than mixtures using overcomplete representations," *IEEE signal processing letters*, vol. 6, no. 4, pp. 87–90, 1999.

[3] Seungjin Choi and Andrzej Cichocki, "Adaptive blind separation of speech signals: Cocktail party problem," in *International conference on speech processing*, 1997, pp. 617–622.

[4] Michael Zibulevsky and Barak A Pearlmutter, "Blind source separation by sparse decomposition in a signal dictionary," *Neural computation*, vol. 13, no. 4, pp. 863–882, 2001.

[5] M Kumar and VE Jayanthi, "Blind source separation using kurtosis, negentropy and maximum likelihood functions," *International Journal of Speech Technology*, vol. 23, no. 1, pp. 13–21, 2020.

[6] Kouhei Sekiguchi, Yoshiaki Bando, Aditya Arie Nugraha, Kazuyoshi Yoshii, and Tatsuya Kawahara, "Fast multichannel nonnegative matrix factorization with directivity-aware jointly-diagonalizable spatial covariance matrices for blind source separation," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 28, pp. 2610–2625, 2020.

[7] Robin Scheibler and Nobutaka Ono, "Fast and stable blind source separation with rank-1 updates," in *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2020, pp. 236–240.

[8] Valentin Leplat, Nicolas Gillis, and Andersen MS Ang, "Blind audio source separation with minimum-volume beta-divergence nmf," *IEEE Transactions on Signal Processing*, vol. 68, pp. 3400–3410, 2020.

[9] Lianru Gao, Zhicheng Wang, Lina Zhuang, Haoyang Yu, Bing Zhang, and Jocelyn Chanussot, "Using low-rank representation of abundance maps and nonnegative tensor factorization for hyperspectral nonlinear unmixing," *IEEE Transactions on Geoscience and Remote Sensing*, 2021.

[10] Christopher A Metzler and Gordon Wetzstein, "Deep s 3 pr: Simultaneous source separation and phase retrieval using deep generative models," in *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2021, pp. 1370–1374.

[11] Pau Bofill and Michael Zibulevsky, "Underdetermined blind source separation using sparse representations," *Signal processing*, vol. 81, no. 11, pp. 2353–2362, 2001.

[12] Lan Ma, Thierry Blu, and William SY Wang, "Event-related potentials source separation based on a weak exclusion principle," in *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*. IEEE, 2017, pp. 1011–1014.

[13] Pierre Comon and Christian Jutten, *Handbook of Blind Source Separation: Independent component analysis and applications*, Academic press, 2010.

[14] Tales Imbiriba, Ricardo Augusto Borsoi, and José Carlos Moreira Bermudez, "Generalized linear mixing model accounting for endmember variability," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2018, pp. 1862–1866.

[15] Rob Heylen, Dževdet Burazerovic, and Paul Scheunders, "Fully constrained least squares spectral unmixing by simplex projection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 11, pp. 4112–4122, 2011.

[16] Lina Zhuang, Chia-Hsiang Lin, Mario AT Figueiredo, and Jose M Bioucas-Dias, "Regularization parameter selection in minimum volume hyperspectral unmixing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 12, pp. 9858–9877, 2019.

[17] Toshihisa Tanaka and Andrzej Cichocki, "Subband decomposition independent component analysis and new performance criteria," in *2004 IEEE International Conference on Acoustics, Speech, and Signal Processing*. IEEE, 2004, vol. 5, pp. V–541.

[18] Tianfeng Chai and Roland R Draxler, "Root mean square error (rmse) or mean absolute error (mae)?–arguments against avoiding rmse in the literature," *Geoscientific model development*, vol. 7, no. 3, pp. 1247–1250, 2014.

[19] Feiyun Zhu, Ying Wang, Bin Fan, Shiming Xiang, Geofeng Meng, and Chunhong Pan, "Spectral unmixing via data-guided sparsity," *IEEE Transactions on Image Processing*, vol. 23, no. 12, pp. 5412–5427, 2014.