# UNDERDETERMINED SOURCE SEPARATION OF EEG SIGNALS IN THE TIME-FREQUENCY DOMAIN

*Zeyong Shan, Jacob Swary, and Selin Aviyente*

Department of Electrical and Computer Engineering
Michigan State University, East Lansing, MI 48824, USA
{shanzeyo, swaryjac, aviyente@egr.msu.edu}

## ABSTRACT

Human brain activity can be measured with high temporal resolution by recording the electric potentials on the scalp surface using imaging methods such as the electroencephalogram (EEG). The analysis of EEG data is difficult due to the fact that multiple neurons may be simultaneously active and the potentials from these sources are superimposed on the limited sensors. It is desirable to unmix the data into signals representing the behavior of the original individual neurons. This is a problem of underdetermined blind source separation (UBSS). Since EEG signals are non-stationary, in this paper a two-stage UBSS approach is proposed for the separation of EEG signals by taking advantage of the high resolution of time-frequency distributions. Experimental results indicate the effectiveness of the introduced approach at separating EEG signals in the time-frequency domain compared with independent component analysis (ICA).

***Index Terms***— Electroencephalogram, blind source separation, time-frequency distribution, sparsity

## 1. INTRODUCTION

The electroencephalogram (EEG) is a recording of the electrical potentials on the scalp, revealing the electrical activity of the brain tissue. Since there are many simultaneous sources of electrical activity in the brain, the currents generated by all of these sources and propagating throughout the head are linearly superimposed on one another. Thus, any electrode will pick up a linear combination of all of the active sources. It is highly desirable to unmix the recorded data into signals representing the behaviors of the original individual sources. This problem is generally called blind source separation (BSS).

In the last several years, there has been much work on the problem of blind source separation, which has resulted in many diverse approaches (e.g. [1]). A key assumption of these algorithms is the statistical independence of the source signals, which results in independent component analysis (ICA). ICA has been used extensively in analysis of brain imaging data including EEG [2]. One condition to apply the ICA technique to source separation is that the number of independent signal sources is the same as the number of sensors. However, since we usually do not know the effective number of statistically independent brain signals contributing to the EEG recorded from the scalp, ICA may not be ideally suitable for performing source separation on EEG signals.

Since there are a limited number of electrode measurements, source separation based on EEG signals is an underdetermined problem. Underdetermined blind source separation (UBSS) is a challenging problem since the mixing matrix is not invertible. Therefore, additional *a priori* information about the sources is needed to allow for reconstruction. One increasingly popular and powerful assumption is the sparsity of the sources for a given basis [3]. Sparse signal representations lend themselves to good separability of the sources, because most of the energy of a basis coefficient at any time instant belongs to a single source.

Since EEG signals are non-stationary and not sufficiently sparse in the time domain, the time-frequency representations of underlying signals are used for source separation. Time-frequency analysis results in high resolution and sparse representations of the mixtures. In this paper, we use a two-stage sparse representation approach for source separation on EEG signals in the time-frequency domain, and compare its performance with that of ICA.

## 2. BACKGROUND ON TIME-FREQUENCY ANALYSIS

A time-frequency distribution (TFD), $X(t,\omega)$, from Cohen's class can be expressed as [1] [4]:

$$X(t,\omega) = \int\int\int \phi(\theta,\tau) s(u+\frac{\tau}{2}) s^*(u-\frac{\tau}{2}) e^{j(\theta u - \theta t - \omega\tau)} du\, d\theta\, d\tau, \quad (1)$$

where $\phi(\theta,\tau)$ is called the kernel function and $s(t)$ is the signal. In this paper, we will use reduced interference distribution (RID) kernels to minimize the effect of cross terms and increase the sparsity level.

[1] All integrals are from $-\infty$ to $\infty$ unless otherwise stated.

## 3. TIME-FREQUENCY BASED SPARSE REPRESENTATION APPROACH FOR UBSS

In this section, a two-stage approach for the UBSS problem in the time-frequency domain is presented, in which the first stage is for determining the mixing matrix, and the second stage is for estimating the source signals.

### 3.1. Linear Mixture Model and Assumptions

Assume that the $M$ mixtures, $\mathbf{z}(t) = [z_1(t), z_2(t), \ldots, z_M(t)]^T$, of the $N$ non-stationary complex source signals, $\mathbf{s}(t) = [s_1(t), s_2(t), \ldots, s_N(t)]^T$, are given by $\mathbf{z}(t) = \mathbf{B}\mathbf{s}(t)$, where $\mathbf{B}$ is the $M \times N$ instantaneous mixing matrix ($M < N$).

Each mixture, $z_i(t)$, is first transformed to the time-frequency plane, and then the corresponding time-frequency distribution is vectorized to form a matrix of time-frequency distributions, $\mathbf{X}$. In our source separation problem, the observed time-frequency distributions, $\mathbf{X}$, can be written as a linear combination of the original sources' TFDs, $\mathbf{S}$, assuming the cross-terms between the sources are negligible by using a RID: $\mathbf{X} \approx \mathbf{B}^2\mathbf{S} = \mathbf{A}\mathbf{S}$, where $\mathbf{X} = [\mathbf{x}_1, \cdots, \mathbf{x}_P] \in R^{M \times P}$ is the mixtures of the sources, $P$ is the total number of time and frequency points, $\mathbf{A} = \mathbf{B}^2 = [\mathbf{a}_1, \cdots, \mathbf{a}_N] \in R^{M \times N}$ is an unknown mixing matrix, $\mathbf{B}^2$ is the element-by-element square of the mixing matrix in the time domain, and $\mathbf{S} = [\mathbf{s}_1, \cdots, \mathbf{s}_P] \in R^{N \times P}$ is the time-frequency representations of the $N$ unknown source signals.

### 3.2. Determination of the Mixing Matrix

Based on the assumption that the source signals are sparse in the time-frequency domain, there exists many columns of $\mathbf{S}$ with only one nonzero entry. For instance, suppose that $\mathbf{s}_{i_1}, \cdots, \mathbf{s}_{i_K}$ are $K$ columns of $\mathbf{S}$, where only the first entry of each of these columns is nonzero, then we have

$$\mathbf{A}\mathbf{s}_{i_j} = \mathbf{a}_1 s_{1i_j} \quad j = 1, \cdots, K, \tag{2}$$

and

$$[\mathbf{x}_{i_1}, \cdots, \mathbf{x}_{i_K}] = \mathbf{A}[\mathbf{s}_{i_1}, \cdots, \mathbf{s}_{i_K}] = [\mathbf{a}_1 s_{1i_1}, \cdots, \mathbf{a}_1 s_{1i_K}], \tag{3}$$

where, $\mathbf{x}_{i_j}$ is the $i_j$th column of $\mathbf{X}$, $\mathbf{a}_1$ is the first column of $\mathbf{A}$, and $s_{1i_j}$ is the first entry of $\mathbf{s}_{i_j}$. From equation (3), we see that each $\mathbf{x}_{i_j}$ is equal to $\mathbf{a}_1$ multiplied by a scalar $s_{1i_j}$, which means that these $K$ column vectors of $\mathbf{X}$, $\mathbf{x}_{i_1}, \cdots, \mathbf{x}_{i_K}$, are distributed along the direction of $\mathbf{a}_1$. Thus, ideally after normalization, $\mathbf{x}_{i_1}, \cdots, \mathbf{x}_{i_K}$ are mapped to a unique vector on the multidimensional unit circle which is equal to $\mathbf{a}_1$. However, in practice, since the mixture matrix $\mathbf{X}$ is not completely sparse in the time-frequency domain, $\mathbf{x}_{i_1}, \cdots, \mathbf{x}_{i_K}$ are not exactly in the same direction as $\mathbf{a}_1$, but rather in the neighborhood of $\mathbf{a}_1$. This means that $\mathbf{a}_1$ lies at the center of $\mathbf{x}_{i_1}, \cdots, \mathbf{x}_{i_K}$. Therefore, we use the $K$-means clustering method to cluster the column vectors of the mixture matrix $\mathbf{X}$ into different clusters, where the center of each cluster corresponds to one column vector of the mixing matrix $\mathbf{A}$. By doing so, we can obtain an estimate of the mixing matrix $\mathbf{A}$.

### 3.3. Estimation of the Source Signals for a Given Mixing Matrix

After obtaining the estimated mixing matrix, the next stage is to estimate the source signals. For a given mixing matrix $\mathbf{A}$, the source signals can be estimated by maximizing the posterior distribution $P(\mathbf{S}|\mathbf{X}, \mathbf{A})$ of $\mathbf{S}$. Under the assumption that the prior is Laplacian, maximizing posterior distribution can be implemented by solving the following optimization problem [5]:

$$\min \sum_{i=1}^{N} \sum_{j=1}^{P} |s_{ij}|, \quad \text{subject to } \mathbf{A}\mathbf{S} = \mathbf{X}. \tag{4}$$

It is not difficult to prove that the optimization problem (4) is equivalent to the following set of $P$ smaller scale linear programming (LP) problems:

$$\min \sum_{i=1}^{N} |s_{ij}|, \quad \text{subject to } \mathbf{A}\mathbf{s}_j = \mathbf{x}_j \quad \text{for } j = 1, \cdots, P. \tag{5}$$

## 4. EXPERIMENTAL RESULTS ON EEG SIGNALS AND PERFORMANCE ANALYSIS

In this section, the presented approach is applied to EEG signals to illustrate its effectiveness to extract sparse source signals in the time-frequency domain from a few number of mixtures. The binomial kernel [4] is used for computing the TFD with 64 time points and 33 frequency points. The results are compared to ICA since it is widely used in the analysis of EEG signals.

### 4.1. EEG Data Set

The EEG data used in this paper is released by Swartz Center for Computational Neuroscience at the University of California, San Diego [6]. During a selective visual attention experiment, stimuli appeared briefly in any of five squares arrayed horizontally above a central fixation cross. In each experimental block, one (target) box was differently colored from the rest. Whenever a square appeared in the target box, the subject was asked to respond quickly with a right thumb button press. The data were constructed by concatenating three-second epochs from one subject, each containing a target square in the attended location (left locations 1 or 2 only). The stimulus was repeated 40 times at the two attended locations resulting in a total of 80 trials per electrode for 32

electrodes. The sampling rate was 128 Hz. Data from 6 electrodes, F3, F4, Cz, P3, P4, and Oz, are chosen to be used in our analysis.

### 4.2. Component Extraction

We first do the source extraction at the single-trial level. Single-trial source separation accounts for the variability across trials which is not possible with the average EEG. Since 6 electrodes are used, there are 6 mixtures of sources per trial. For single-trial source extraction, we first assume the number of sources to be 8. This number is chosen since higher number of sources resulted in sources that were too sparse and did not correspond to actual neuronal activity. ICA is applied to the same data and can only extract 6 sources from the 6 mixtures. Both of these BSS techniques are applied to all 80 trials of data. This decomposition results in 640 and 480 single-trial components extracted by the proposed method and ICA, respectively. Since it is difficult to do a one-to-one comparison between the extracted components, we propose to use the following data reduction scheme for evaluation purposes.

In order to evaluate the performance of the proposed method and ICA, the single-trial results are put together in their respective groups depending on stimulus type. The 8 extracted components of trial $v$ are represented by the matrix, $\mathbf{S}_v$, of size $8 \times P$, where $P$ is the total number of time and frequency points and in our case is equal to 2112. For the data reduction of the extracted components for a particular stimulus, the extracted components over all trials are each appended to form a new matrix, $\tilde{\mathbf{S}}_u$, such that

$$\tilde{\mathbf{S}}_u = \begin{bmatrix} \mathbf{S}_1 \\ \mathbf{S}_2 \\ \vdots \\ \mathbf{S}_{40} \end{bmatrix} = \begin{bmatrix} s_1^1(1) & \cdots & s_1^1(2112) \\ \vdots & \vdots & \vdots \\ s_8^1(1) & \cdots & s_8^1(2112) \\ s_1^2(1) & \cdots & s_1^2(2112) \\ \vdots & \vdots & \vdots \\ s_8^{40}(1) & \cdots & s_8^{40}(2112) \end{bmatrix}, \quad (6)$$

where $u = \{1, 2\}$ represents the locations of the stimulus, and $\tilde{\mathbf{S}}_u$ is of size $320 \times 2112$. Each element, $s_i^v(j)$, is one time-frequency point of source $i$ from trial $v$.

$K$-means clustering is then carried out on each $\tilde{\mathbf{S}}_u$ where each of its rows is grouped into one of $K$ clusters based on the Euclidean distance to the cluster center. The clustering algorithm is run 10 times to avoid randomness in the final cluster results, and the algorithm is run for different values of $K$ at 8, 12, and 16 to determine how the number of clusters affects the outcome. The rows of $\tilde{\mathbf{S}}_u$ are grouped based on the results of the 10 $K$-means runs to create a matrix, $\mathbf{R}$, of size $320 \times 320$. The entry, $r_{ij}$, represents how many times, out of 10, row $i$ of $\tilde{\mathbf{S}}_u$ is grouped into the same cluster as row $j$ of $\tilde{\mathbf{S}}_u$. This matrix serves as a similarity measure, the more times two sources are grouped together by $K$-means, the more similar they are. A hierarchical clustering is then carried out using the similarity matrix, $\mathbf{R}$, as its distance metric to generate $K$ clusters. Cluster centers are calculated by the mean of the time-frequency components in each cluster, and these are the components that categorize all single-trial EEG results. For example, a set of extracted components for $K = 8$ are shown in Figs. 1 and 2 for ICA and the proposed UBSS method, respectively.

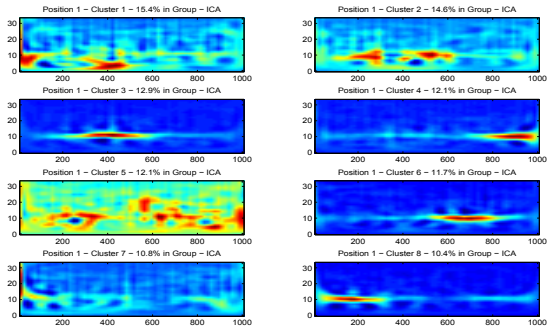


**Fig. 1**. Components extracted using ICA over all single-trial results for stimulus location $u = 1$

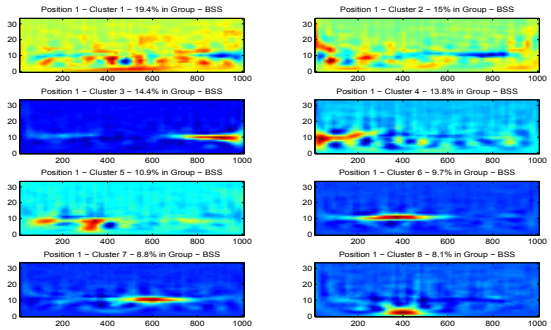


**Fig. 2**. Components extracted using the proposed UBSS method over all single-trial results for stimulus location $u = 1$

### 4.3. Performance Evaluation

The levels to which the extracted components are sparse, disjoint, and localized in the time-frequency domain all speak to how close they may be to an actual underlying source in the brain. The components obtained from the clustering method described in the previous section are evaluated based on these factors.

Since a sparse component must have most of its values close to zero, the $l_1$-norm is a good measurement of how sparse a component is. The extracted clusters are represented by the $K \times 2112$ matrix $\mathbf{C}_u, u = \{1, 2\}$ and each component's sparsity is measured with

$$\sum_{m=1}^{2112} |c_i^u(m)|, \quad (7)$$

where $u$ refers to stimulus location, and $i$ represents component number between 1 and $K$.

Disjointness between two components is measured by using the inner product. A summation of all the pairwise inner products between $K$ components represents a total level of disjointness over all extracted components and is computed by

$$\sum_{i \neq j} \sum_{m=1}^{2112} |c_i^u(m) c_j^u(m)|. \quad (8)$$

Time-frequency localization of each component is computed using a measurement of entropy. This is calculated as

$$- \sum_{m=1}^{2112} |c_i^u(m)| \log_2 |c_i^u(m)|. \quad (9)$$

A smaller entropy value corresponds to a more localized component.

The results comparing the proposed UBSS method with ICA based on these parameters are shown in Tables 1, 2, and 3. Under the proposed approach, the extracted components are typically more sparse, localized, and disjoint than the extracted components under ICA. This means that under the proposed approach, the components are more likely a closer representation of true sources. In addition, as the value of $K$ increases, the extracted components are less disjoint from each other. This shows that some components belonging to the same source are considered as different sources. Thus, it is important to correctly choose the number of clusters. Finally, the extracted components from both methods are projected back to the electrodes to show the amount of variance of the original data explained by the sources. From Table 4, we can see that almost the same amount of variance is explained by the components extracted from both methods.

**Table 1**. Mean measure of $l_1$ norm to show sparsity

| Running Conditions | Location 1 (u=1) | | Location 2 (u=2) | |
|---|---|---|---|---|
| | UBSS | ICA | UBSS | ICA |
| *K*=8 | 23.03 | 29.06 | 21.92 | 27.63 |
| *K*=12 | 22.36 | 28.31 | 22.54 | 28.15 |
| *K*=16 | 23.29 | 28.18 | 22.43 | 27.27 |

**Table 2**. Mean measure of entropy to show time-frequency localization

| Running Conditions | Location 1 (u=1) | | Location 2 (u=2) | |
|---|---|---|---|---|
| | UBSS | ICA | UBSS | ICA |
| *K*=8 | 9.80 | 10.29 | 9.73 | 10.25 |
| *K*=12 | 9.79 | 10.25 | 9.80 | 10.26 |
| *K*=16 | 9.85 | 10.24 | 9.81 | 10.20 |

**Table 3**. Measure of disjointness by correlation between components

| Running Conditions | Location 1 (u=1) | | Location 2 (u=2) | |
|---|---|---|---|---|
| | UBSS | ICA | UBSS | ICA |
| *K*=8 | 3.12 | 3.63 | 3.84 | 4.35 |
| *K*=12 | 8.92 | 9.38 | 8.30 | 8.90 |
| *K*=16 | 14.05 | 14.51 | 15.87 | 16.35 |

**Table 4**. Average component projection to electrodes

| Running Conditions | Location 1 (u=1) | | Location 2 (u=2) | |
|---|---|---|---|---|
| | UBSS | ICA | UBSS | ICA |
| *K*=8 | 0.026 | 0.039 | 0.029 | 0.045 |
| *K*=12 | 0.058 | 0.091 | 0.065 | 0.103 |
| *K*=16 | 0.129 | 0.217 | 0.147 | 0.246 |

## 5. CONCLUSIONS

This paper introduces a two-stage approach for underdetermined blind separation of sparse and non-stationary sources using TFDs. The performance of the proposed method is compared to ICA by applying both algorithms to the multiple trial EEG data set. Data reduction by clustering is performed over all single-trial results to extract components that represent the sources. The presented approach provides components that are more sparse and localized in the time-frequency domain and that are more distinct from each other than ICA. The proposed method can also be used as an effective data reduction method.

## 6. REFERENCES


[1] A. J. Bell and T. J. Sejnowski, "An information–maximization approach to blind separation and blind deconvolution," *Neural Computation*, vol. 7, pp. 1129–1159, 1995.

[2] S. Makeig, A.J. Bell, T.-P. Jung, and T.J. Sejnowski, "Independent component analysis of electroencephalographic data," *Advances in Neural Information Processing Systems*, 1996.

[3] S. Mallat and Z. Zhang, "Matching pursuits with time–frequency dictionaries," *IEEE Trans. on Signal Processing*, vol. 41, pp. 3397–3415, 1993.

[4] L. Cohen, *Time–Frequency Analysis*, Prentice Hall, New Jersey, 1995.

[5] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Scientific Comp.*, vol. 20, pp. 33–61, 1999.

[6] *http://www.sccn.ucsd.edu/eeglab*.