Blind Separation of Digital Sources

Yuanqing Li†, Andrzej Cichocki and Liqing Zhang‡

†Automatic Control Engineering Department
South China University of Technology
Guangzhou, 510641, China

‡Laboratory for Advanced Brain Signal Processing
RIKEN Brain Science Institute, Wako shi, Saitama, 3510198, Japan

Abstract: This paper presents novel techniques for blind separation of instantaneously mixed digital sources, which are suitable for the case with less sensors than sources. First, a solvability analysis is presented for a general case. A necessary and sufficient condition for recoverability of sources is derived. Then the binary source case is considered, a new deterministic blind separation algorithm is proposed to estimate the mixing matrix and separate all sources efficiently in the noise-free or low noise case. By estimating the probability density function of the mixtures, the deterministic algorithm is extended to deal with the case of high noise. Finally, simulation results are presented to illustrate the validity of the algorithms.

Keywords: Digital source, blind separation, solvability conditions.

1 Introduction

Digital signals play important roles in pattern recognition, digital signal processing. When multiple digital signals are transmitted from different sources, the mixtures of them are often received by sensors. In this paper we consider a linear instantaneous mixing model described as,

\[ \mathbf{x}(k) = \mathbf{A}\mathbf{s}(k) + \mathbf{v}(k), \]

where \( \mathbf{s}(k) = [s_1(k), \ldots, s_n(k)]^T \) is an unknown \( n \)-dimensional vector of mutually independent digital source signals which can take only \( J \) discrete values \{\( d_1, d_2, \ldots, d_J \)\} denoted as \( D \), e.g., \{0, 1\} or \{-1, +1\} in binary case. \( \mathbf{x}(k) = [x_1(k), \ldots, x_m(k)]^T \) is an available \( m \)-dimensional sensor (mixed) signal vector, and \( \mathbf{A} \) is an unknown \( m \times n \) constant mixing matrix. \( \mathbf{v}(k) \in \mathbb{R}^m \) is the additive white Gaussian noise vector with zero-mean.

In (1), the sources and the mixing matrix are unknown, only the sensor signals \( \mathbf{x}(k) \) are observed. The task of blind separation is to recover sources \( s_1, \ldots, s_n \) from \( x_1, \ldots, x_m \).

Previous related works include the geometrical approach [2, 3], Maximum Likelihood Approach [9], [10], iterative algorithm [1], [4], [11], matrix factorization [5], [8], deterministic approach [7], etc. The geometrical approach is efficient only in the very low noise case. The maximum
likelihood algorithm often depends on initial values and suffers from wrong convergence. In iterative algorithm, the choice of starting point and algorithm convergence are key problems which limit it’s application as in ML approach. In [5],[8], the mixing matrix $A$ was assumed to be full of column rank as in [6]. In [7], a deterministic approach was proposed also under the condition of low noise or no noise as in [3]. Additionally, the solvability was not discussed in all references above.

This paper discusses blind separation of digital sources. The solvability analysis is presented in Section 2. Blind separation algorithm follows in Sections 3. Simulation results are presented in Section 4. The concluding remarks in Section 5 review the advantages of the proposed approach and states the main limitations.

2 Solvability Analysis

This section analyzes solvability for blind separation of digital sources. Rewrite mixing matrix $A = [a_1, a_2, \cdots, a_n]$, where $a_k$’s are nonzero column vectors of $A$, then the corresponding noise free model of (1) can be represented as follows:

$$x(k) = a_1 s_1(k) + \cdots + a_n s_n(k).$$

(2)

At first, we introduce two definitions.

Definition 1 The model (2) is said to be well-posed, if and only if $As = As'$ implies that $s = s'$, where $s, s'$ are two digital source vectors with components valued in $\{d_1, d_2, \cdots, d_J\}$.

Denote the set $D_1 = \{d_i - d_j | d_i, d_j \in D, i \neq j\}$. Notice that $D_1 = \{-1, 0, 1\}$ as $D = \{0, 1\}$. We have the following theorem:

Theorem 1 System (2) is well-posed, if and only if column vectors of $A$ satisfy the following inequality:

$$c_1 a_{i_1} + \cdots + c_L a_{i_L} \neq 0,$$

(3)

for any $c_k \in D_1, k = 1, \cdots, L$, and $\{i_1, \cdots, i_L\}$ covers all subsets of $\{1, 2, \cdots, n\}$.

The proof is is omitted here.

3 Blind Separation Algorithm

The algorithm in this section is for binary sources.

The low noise case including noise free case is considered firstly, and a deterministic blind separation algorithm is proposed. For high noise case, the deterministic algorithm is also efficient but then the cluster centers must be estimated correctly by other methods (e.g., pdf estimation in this paper) in advance.

Suppose that the conditions of Theorem 1 are always satisfied in this section. Since sources are binary, there are $2^n$ different output vectors of the noise free model (2), denoted as $\{s_1, \cdots, s_N\}$, where $N = 2^n$. Set

$$d = \min \{|x_i - x_j| \natural i, j = 1, \cdots, N, i \neq j\}.$$  

(4)

The parameter $d$ is an important factor in analyzing noise tolerance for binary sources separation. It is larger $d$ not larger source energy that implies larger noise tolerance. Thus in this paper, the following parameter is used instead of signal-noise ratio $SNR$,

$$SDR = \frac{\sigma^2}{d^2},$$

(5)

where $\sigma^2$ is the sum of variance of all noise components.

3.1 Low noise or noise free case

At first, we present an assumption on low noise.

Assumption 1: The noise vector in (1) satisfies the following condition:

$$\|n(k)\| < \frac{d}{4}.$$  

(6)
Obviously, under the condition (6), there are $N = 2^n$ different clusters with radius less than $\frac{d}{2}$ formed by the outputs of (1) represented by $N$ vectors $\{x_1, \ldots, x_N\}$, which are the center vectors of the clusters. For convenience, rewrite these vectors as a matrix,

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1N} \\ x_{21} & x_{22} & \cdots & x_{2N} \\
\vdots & \vdots & \ddots & \vdots \\
x_{m1} & x_{m2} & \cdots & x_{mN} \end{bmatrix}.$$  

Remarks 1: In this section, two mixture vectors $x(k_1)$ and $x(k_2)$ are said to be in the same cluster if

$$||x(k_1) - x(k_2)|| < \frac{d}{2}. \quad (7)$$

Now we present the blind separation algorithm by which the columns of mixing matrix $A$ are estimated, each of the cluster centers are labeled as a column vector $[s_1, \ldots, s_n]$, where $s_i \in \{d_1, d_2\}$, and all sources are recovered.

(The first grouping)

1. Choose a row of the matrix $X$ with at least two non-zero components assumed to be the first row, and then find out the biggest and the second biggest components assumed to be $x_{11}$ and $x_{12}$. Set

$$\hat{a}_1 = \frac{1}{d_2 - d_1} [x_{11} - x_{12}].$$

It is not difficult to prove that $\hat{a}_1$ is one of columns of $A$ up to a sign (see Appendix 1).

2. Given $\hat{a}_1$, we match the columns of $X$ pairwisely. That is, if $||x_i - x_j - (d_2 - d_1)\hat{a}_1|| < \epsilon_0$, where $\epsilon_0$ is a sufficiently small constant chosen in advance, then $(x_i, x_j)$ is defined as a pair. It can be proved that under Assumption 2, for any $x_i$, there exists only one matched cluster center (see Appendix 2).

There exist $\frac{N}{2}$ pairs denoted as $(x_{i_1}, x_{i_2}), \ldots, (x_{i_{(N-1)/2}}, x_{i_{N/2}})$.

3. According to the pairs above, divide $\{x_i\}$ into two groups denoted as two matrices,

$$X^{11} = [x_{i_1}, x_{i_3}, \ldots, x_{i_{(N-1)/2}}],$$

$$X^{12} = [x_{i_2}, x_{i_4}, \ldots, x_{i_{N/2}}].$$

4. For all columns of $X^{11}$, set the first components of their labeling vectors to be $d_2$, and for all columns of $X^{12}$, set the first components of their labeling vectors to be $d_1$.

5. Repeating the above processes 1, 2 and 3 by replacing $X$ with $X^{11}$, we obtain another column of $A$ up to a sign denoted as $\hat{a}_2$, and two groups denoted as $X^{21}$ and $X^{22}$. For all columns of $X^{21}$, set the second components of their labeling vectors to $d_2$, and set the second components of the labeling vectors to $d_1$ for any cluster centers of $X^{22}$.

6. Using $\hat{a}_2$, we divide all columns of $X^{12}$ into two groups denoted as $X^{23}$ and $X^{24}$. For all columns of $X^{23}$, set the second components of their labeling vectors to $d_2$, and set the second components of the labeling vectors to $d_1$ for any cluster centers of $X^{24}$.

Repeating the above process, we can divide each of $X^{21}$, $X^{22}$, $X^{23}$ and $X^{24}$ into two groups, obtain another column of $A$ up to a sign and the third components of all labeling vectors. After the $n-th$ grouping, we can obtain all columns of $A$ denoted as $[\hat{a}_1, \ldots, \hat{a}_n]$ up to a sign and order. Also we can obtain the labeling vectors of all cluster centers.

(Separation)

Given a mixture vector $x(k)$, find out the closest cluster center, then the corresponding labeling vector $\hat{s}(k)$ is the source vector $s(k)$ up to a exchange of $d_1$ and $d_2$, and source order, $k = 1, 2, \ldots$.

3.2 High noise case

It is not difficult to find that if the cluster centers can be estimated correctly in advance, then the proposed deterministic
algorithm still can be used. Thus the separation strategy for high noise case is divided into two steps, the first is to estimate cluster centers, the second is to separate sources as in Subsection 3.1. The main task of this subsection is to propose a method of estimating the cluster centers.

It is noted that the sample points of zero mean Gaussian noise distribute almost near the zero. For instance, for one dimensional standard normal distribution, the probability of the affair $\{ |v| > 3 \}$ is less than 0.05, where $v$ is sample point. Thus for model (1), all output vectors produce $N(=2^n)$ clusters which are overlapped partially when the noise is not too large. And the probability density function of the output vectors of (1) has maximum in the centers of the clusters as illustrated in the first subplot of Fig. 1 on one dimensional mixture.

Now we present the algorithm for estimating the cluster centers. Without loss of generality, consider the model (1) with one dimensional output.

1. Estimate the pdf of the output $x$.

Suppose that there are $N_0$ sample points of $x$ denoted as a set $X$, where $N_0 \gg N$, and that the minimum and maximum in $X$ are assumed to be $a$, $b$ respectively. The interval $[a, b]$ is then divided equally into $M$ sub-intervals which are $[a + i \delta, a + (i + 1) \delta]$, $i = 0, \cdots, M - 2$, and $[a + (M - 1) \delta, b]$, where $\delta = \frac{b - a}{M}$, and $M$ is a sufficiently large positive integer. By estimating the number of sample points in each interval denoted by $n_i$ for the $i-th$ interval, the probability for $x$ belonging to the $i-th$ interval can be obtained, that is, $p'_i = \frac{n_i}{N_0}$, $i = 1, \cdots, M$. The first subplot in Figure 1 shows the probability density function with one mixture of three text image sources and Gaussian noise ($SND = 12.9$dB, $M = 300$), where the three sources can be seen in Example 2.

2. Find out the $N$ peaks of the pdf.

At first, the pdf is smoothed by the following filter (several smoothings may be more useful sometimes),

$$p_k = \frac{1}{16} [p'_{k-2} + 4p'_{k-1} + 6p'_k + 4p'_{k+1} + p'_{k+2}],$$

$$\text{(8)}$$

Then the following criterion is used for finding out the $N$ peaks of the smoothed pdf assumed to be $p_1, \cdots, p_N$.

**Criterion:** If the three conditions are satisfied, (1) $p_i + p_{i+1} + p_{i+2} \leq p_{i+1} + p_{i+2} + p_{i+3}$; (2) $p_{i+1} + p_{i+2} + p_{i+3} > p_{i+2} + p_{i+3} + p_{i+4}$; (3) $p_{i+1} > \frac{n_i}{N_0}$ or $p_{i+2} > \frac{n_i}{N_0}$, or $p_{i+3} > \frac{n_i}{N_0}$, then $p_{i+2}$ is taken as a peak, where $n_0$ is a positive integer chosen in advance. Additionally, if $p_1 + p_2 > p_2 + p_3 > p_3 + p_4 + p_1 > \frac{n_1}{N_0}$ or $p_2 > \frac{n_1}{N_0}$ or $p_3 > \frac{n_1}{N_0}$, then $p_2$ is said to be a peak. If $P_{M-2} + \sum_{i=1}^{PM-1} + P_M > P_{M-3} + P_{M-2} + P_{M-1}$, $P_{M-2} > \frac{n_i}{N_0}$, or $P_{M-1} > \frac{n_i}{N_0}$ or $P_M > \frac{n_i}{N_0}$, then $P_{M-1}$ is also chosen as a peak.

3. Set $\tilde{x}_1 = a + (i_1 + \frac{1}{2}) \delta$, $\cdots, \tilde{x}_N = a + (i_N + \frac{1}{2}) \delta$, which are the estimated cluster centers. The second subplot in Figure 1 presents the 8 centers corresponding to the first subplot.

After the cluster centers are estimated, we can use the deterministic algorithm in subsection 3.1 for blind separation of the binary sources.

![Figure 1](image-url)
4 Simulation Results

In this section, two simulation examples are presented.

**Example 1:** Consider the following model,

\[
x(i, j) = [3.5, 3.0, 4.2][s_1(i, j), s_2(i, j), s_3(i, j)]^T + v(i, j),
\]

where \(s_1, s_2, s_3\) are three binary text sources with \(250 \times 250\) pixels, \(v\) is Gaussian white noise which satisfies Assumption 1, is about \(25dB\) on \(\hat{d}^2 = 0.25\). Of course, only the mixture \(x\) is available in the blind separation.

Fig. 2 shows the blind separation results, in which, the first subplot is the mixture, the three subplots in the second row are recovered source text images. Since the recovered sources have no obvious difference with the original ones, the original sources are not presented here.

![Figure 2: Blind separation for one mixture of three text sources in Example 1. The first row: available image corrupted by other text images, the second row: three reconstructed texts.](image)

![Figure 3: The curve of bit error rate vs SDR in Example 2.](image)

5 Concluding Remarks

A novel approach for blind separation and extraction of linearly mixed digital sources is proposed in this paper. Necessary and sufficient condition on recoverability of digital sources is obtained. For binary source case, a deterministic blind separation algorithm is proposed for low noise or no noise mixtures, in which the estimating mixing matrix and labeling clusters are carried out simultaneously. For high noise case, the blind separation algorithm also can be used successfully by combining with probability density function estimation. The simulation results confirm the validity of the approach in this paper.

There are several points to be emphasized as the concluding remarks of this paper.

1) The noise tolerance of blind separation algorithm depends on the minimal distance of \(J^n\) different output vectors of
the corresponding noise free model (2).

2) If the number of sources $n$ is not large, e.g., $n \leq 10$ for binary sources, the cluster number is also not large, the algorithms in this paper are very efficient, by which the real-time separation and extraction are possible.

If the source $n$ is very large, it is difficult to obtain all different clusters, the computation time of the algorithms in this paper will increase exponentially. The open problem is how to reduce computational complexity.

References


