BLIND SEPARATION OF N BINARY SOURCES FROM ONE OBSERVATION: A DETERMINISTIC APPROACH

Konstantinos I. Diamantaras

Department of Informatics
Technological Education Institute
of Thessaloniki
GR-54101 Sindos, Greece
kdiamant@it.teithe.gr

Efthymia Chassioti

University of Athens
Department of Informatics
TYPA Buildings, Panepistimioupolis
GR-15784, Athens, Greece

ABSTRACT

We show that it is possible to separate $n > 1$ binary source signals from a single linear mixture, under very mild assumptions, based on the clustering of the data. We develop the mathematical treatment of the problem and propose a recursive, finite algorithm for its solution. The application of this method is only limited by the level of noise and by the combinatorial explosion as the number of sources n increases since the algorithm complexity is exponential w.r.t. n. Simulation results indicate that the method can successfully separate as many as 10 sources from a single mixture in a few seconds on a typical desktop PC.

1. INTRODUCTION

Binary signals are very common in signal processing and digital communications. We consider binary variables taking values in the set $\{-1, 1\}$ (also known as binary *antipodal* variables) but the results can be easily extended in the set $\{0, 1\}$. When multiple signals are transmitted from different sources, the receiver often records a mixture of them. If reverberation or mutlipath issues are ignored then the mixture can be modelled by a linear weighted sum with constant mixing parameters. This model has been extensively used in speech processing, digital communications, and medical signal processing. In the case where neither the mixing parameters nor the sources are known the term "blind" is added on the problem. The fast growing literature on Blind Source Separation (BSS) from linear instantaneous mixtures is usually based on an appropriate statistical problem formulation. Higher order statistics are used in the Independent Component Analysis (ICA) method which assumes that the sources are white, independent random sequences [3, 2]. Second order statistics can be also used provided that the source signals are colored [1].

In this paper we'll study the special case of binary antipodal sources. We'll show that it is possible to blindly separate many such sources from a single linear instantaneous observation using a clustering approach. Clustering based equalizers (CBSE) have been proposed elsewhere [6] but the approach taken was different as it was based on the Viterbi algorithm for trellis code optimization. The separation of more sources than mixtures has been studied before in the context of ICA and maximizing a posteriory probabilities [11]. However, our approach is very different from any standard approach since we shall make no particular assumption on the stochastic properties of the souces; in fact, they may even be deterministic. We'll find that the problem is more combinatorial in nature rather than stochastic. Our approach will lead to a recursive algorithm which is demonstrated to successfully separate 10 sources from a single mixture in a few seconds. The application of this method is only limited by the level of noise and by the combinatorial explosion as the number of sources n increases since the algorithm complexity is exponential with respect to n.

This work was supported by the Greek Secretariat of Research and Technology, Programme PENED, No 99ED83 (ISODIA).

2. PROBLEM FORMULATION

Let $x(k)$ denote the observed one-dimensional sequence, where $k = 1, \cdots, N$, is the time index. The data model is described by a linear mixing vector $\mathbf{a} = [a_1, \cdots, a_n]^T$ and additive noise $e(k)$:

$$x(k) = \mathbf{a}^T \mathbf{s}(k) + e(k) \tag{1}$$

where $\mathbf{s}(k) = [s_1(k), \cdots, s_n(k)]^T$ is the vector source sequence. Each individual source is an i.i.d. binary antipodal sequence: $s_i(k) \in \{-1, +1\}$, for all k, and all individual sources are mutually independent. Furthermore, the vector source sequence $\{\mathbf{s}(k)\}$ is assumed to be rich enough so that every combination vector $[\pm 1, \ \pm 1, \cdots, \pm 1]^T$ appears at least once for some k. We make no other assumptions regarding the probability distribution of the sources, or any statistical property of $\mathbf{s}(k)$. The noise $e(k)$ is assumed to be zero-mean Gaussian with unknown variance σ^2. The problem is to find both the mixing vector $\mathbf{a}$ and the sources $s_i(k)$ upto a permutation of their order and an inherent ambiguity on their sign. The scaling ambiguity, typical of most BSS problems, is not present here since the source samples can take only certain values.

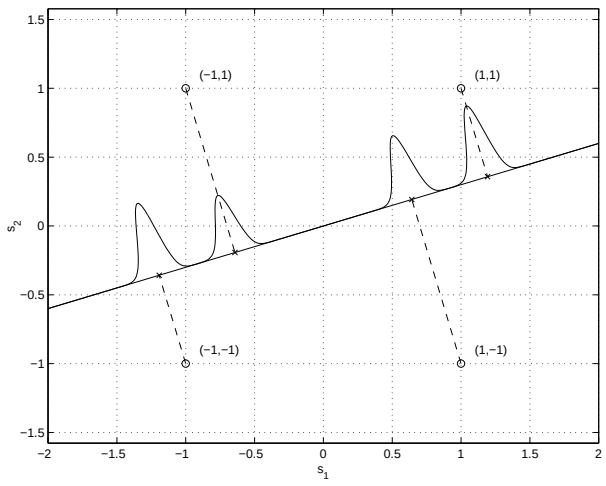

Figure 1: The distribution $P(X)$ of x for two antipodal sources s_1, s_2 is the mixture of 4 Gaussian "bells" of equal variances but different amplitudes. If $\|\mathbf{a}\| = 1$ then the bell centers are the projections of the vectors $[-1,-1]$, $[-1,+1]$, $[+1,-1]$, $[+1,+1]$, on the direction $\mathbf{a}$ as shown in the figure.

3. RETRIEVING THE MIXING VECTOR

Every source vector sample $\mathbf{s}(k)$ can take $M = 2^n$ values –as many as the different combinations of the n binary antipodal variables $s_i(k)$. Let $\mathcal{S} = \{\mathbf{t}_1, \cdots, \mathbf{t}_M\}$ be the set of possible values for $\mathbf{s}(k)$, where $\mathbf{t}_i$, $i = 1, \cdots, M$, are the combination vectors.

The probability distribution of $x(k)$ can be written as follows

$$\begin{aligned} P(x(k) = X) &= \sum_{\mathbf{S} \in \mathcal{S}} P(x(k) = X | \mathbf{s}(k) = \mathbf{S}) P(\mathbf{s}(k) = \mathbf{S}) \\ &= \sum_{i=1}^{M} P(x(k) = X | \mathbf{s}(k) = \mathbf{t}_i) P_i \end{aligned} \quad (2)$$

where P_i stands for $P(\mathbf{t}_i)$.

$$P(x(k) = X) = \frac{1}{\sqrt{2\pi}\sigma} \sum_{i=1}^{M} exp\{-\frac{1}{2\sigma^2}(x(k) - c_i)^2\} P_i \quad (3)$$

$$c_i \stackrel{\triangle}{=} \mathbf{a}^T \mathbf{t}_i \quad (4)$$

From (3) it follows that $P(X)$ is a mixture of $M = 2^n$ Gaussians with the same variance parameter σ^2, with different centers (in general) at $\mathbf{a}^T\mathbf{t}_i$, and with amplitudes P_i. (see Fig. 1).

3.1. Estimating the centers of the mixture model

The estimation of a mixture density from data samples is a classical statistical problem [5][chapter 6],[13]. Especially Guassian mixtures have received a lot of attention due to their applications in pattern recognition and singal processing. Different methods for estimating the parameters in a Gaussian mixture include: Radial Basis Function (RBF) neural networks [12], neural mixtures of experts [9, 10], and the Expectation-Maximization (EM) algorithm [4, 7]. See [8] for a comprehensive treatment of most of these methods. The output of any such technique will be a set of estimates σ^2, $\hat{\mathbf{c}}_i$, $\hat{P}_i$, for the variance, the centers, and the prior probabilities of the mixture density, respectively.

In the noise-free case ($\sigma = 0$) we can use a simple clustering technique, where each sample $x(k)$ forms a new cluster provided that it has distance greater than ϵ from all previous clusters, where ϵ is a small positive number. The cluster centers are the estimates of c_i; the prior probabilities P_i are estimated by the relative frequency of samples in each cluster.

For the perfect centers the following equation holds

$$[c_1 \cdots c_M]^T = [\mathbf{t}_1 \cdots \mathbf{t}_M]^T \mathbf{a} \quad (5)$$

or

$$\mathbf{c} = \mathbf{T}\mathbf{a}$$

where we made the obvious definitions for $\mathbf{c}$, $\mathbf{T}$. I turns out that there is a lot of structure both in the matrix $\mathbf{T}$ and in the vector $\mathbf{c}$, allowing us to find the solution in a finite number of steps as shown next.

3.2. Structure of the center vector

In the following analysis we shall assume that the mixing parameters are positive and arranged in decreasing order, so

$$a_1 > a_2 > \cdots > a_n > 0. \quad (6)$$

This assumption **does not** hurt generality because (a) we can change the sign of any negative mixing parameter a_i if we change, at the same time, the sign of the associated source s_i, (b) we can sort the mixing parameters by their magnitude if we rearrange accordingly the order of the source signals. Since neither the sign nor the order of the source signals are observable, our assumption (6) is most general.

We shall define $\mathbf{T}^{(n)}$, for any n, to be the $2^n \times n$ matrix of n-dim combination vectors $[\pm 1, \cdots, \pm 1]$:

$$\mathbf{T}^{(n)} \stackrel{\triangle}{=} \begin{bmatrix} -1 & -1 & \cdots & -1 & -1 \\ -1 & -1 & \cdots & -1 & 1 \\ -1 & -1 & \cdots & 1 & -1 \\ \vdots & \vdots & & \vdots & \vdots \\ 1 & 1 & \cdots & -1 & 1 \\ 1 & 1 & \cdots & 1 & -1 \\ 1 & 1 & \cdots & 1 & 1 \end{bmatrix}$$

There is an anti-symmetric property of the rows of $\mathbf{T}^{(n)}$: if $\mathbf{t}_i^T$ and $\mathbf{t}_{n-i+1}^T$ are, respectively, the i-th and $(n-i+1)$-th row of $\mathbf{T}^{(n)}$ then

$$\mathbf{t}_i = -\mathbf{t}_{n-i+1}$$

First, we'll show that for $n = 2$ the centers c_i are arranged in increasing order.

Lemma 1 *Let* $\alpha_1 > \alpha_2 > 0$ *and* $[\gamma_1, \cdots, \gamma_4]^T = \mathbf{T}^{(2)}[\alpha_1, \alpha_2]^T$. *Then* $\gamma_1 < \gamma_2 < \gamma_3 < \gamma_4$.

PROOF. It is obvious since

$$\mathbf{T}^{(2)} = \begin{bmatrix} -1 & -1 \\ -1 & 1 \\ 1 & -1 \\ 1 & 1 \end{bmatrix}$$

and so $\gamma_2 - \gamma_1 = 2\alpha_2 > 0$, $\gamma_3 - \gamma_2 = 2(\alpha_1 - \alpha_2) > 0$, $\gamma_4 - \gamma_3 = 2\alpha_2 > 0$, ■

For $n > 2$ we'll show that although the centers c_i may not be arranged in increasing order, the following statements are true:

(a) the sequence c_1, ..., c_M can be partitioned in consecutive quadruples, each arranged in increasing order;

(b) the first 3 centers $c_1 < c_2 < c_3$ are the three smallest values in the sequence c_i;

(c) the last 3 centers $c_{M-2} < c_{M-1} < c_M$ are the three largest values in the sequence c_i;

(d) The two smallest mixing parameters a_{n-1}, a_n, can be retrieved from c_1, c_2, and c_3;

(e) The differences $c_{4i+j} - c_{4i+1}$, $j = 2, 3, 4$, in the i-th quadruple are functions of a_{n-1}, a_n and are independent of i.

Lemma 2 *Let $n > 2$, $M = 2^n$, $\boldsymbol{\gamma} = [\gamma_1, \cdots, \gamma_M]^T = \mathbf{T}^{(n)}\boldsymbol{\alpha}$, where the elements α_1, ..., α_n, of the vector $\boldsymbol{\alpha}$ are positive reals, arranged in decreasing order. Let $\overline{\gamma}_1$, ..., $\overline{\gamma}_M$ be the sequence γ_i arranged in increasing order. Then*

$$\begin{aligned} \overline{\gamma}_1 = \gamma_1 &= -\gamma_M &= -\overline{\gamma}_M \\ \overline{\gamma}_2 = \gamma_2 &= -\gamma_{M-1} &= -\overline{\gamma}_{M-1} \\ \overline{\gamma}_3 = \gamma_3 &= -\gamma_{M-2} &= -\overline{\gamma}_{M-2}\ . \end{aligned}$$

Furthermore,

$$\begin{aligned} \alpha_n &= (\overline{\gamma}_2 - \overline{\gamma}_1)/2 \\ \alpha_{n-1} &= (\overline{\gamma}_3 - \overline{\gamma}_1)/2\ . \end{aligned}$$

PROOF. By definition we have,

$$\begin{aligned} \gamma_1 &= [-1 \cdots -1\ -1\ -1][\alpha_1 \cdots \alpha_{n-2}\ \alpha_{n-1}\ \alpha_n]^T \\ \gamma_2 &= [-1 \cdots -1\ -1\ +1][\alpha_1 \cdots \alpha_{n-2}\ \alpha_{n-1}\ \alpha_n]^T \\ \gamma_3 &= [-1 \cdots -1\ +1\ -1][\alpha_1 \cdots \alpha_{n-2}\ \alpha_{n-1}\ \alpha_n]^T \end{aligned}$$

First we prove that γ_1 is the smallest element in the sequence $\{\gamma_i\}$. Indeed, $\gamma_1 = -\sum_{i=1}^n \alpha_i$ and so for all $j > 1$:

$$\gamma_j - \gamma_1 = \sum_{i=1}^{n} (t_{j,i} + 1)\alpha_i > 0$$

since $\alpha_i > 0$, $t_{j,i} \geq -1$ for all i, and $t_{j,m} = 1 > -1$ for at least one m.

Next we show that γ_2 is the second last element of the sequence $\{\gamma_i\}$. Clearly $\gamma_2 > \gamma_1$. We need then to show that $\gamma_j > \gamma_2$, for all $j > 2$. We have $\gamma_2 = -\sum_{i=1}^{n-1} \alpha_i + \alpha_n$, and so

$$\gamma_j - \gamma_2 = \sum_{i=1}^{n-1} (t_{j,i} + 1)\alpha_i + (t_{j,n} - 1)\alpha_n$$

The number $t_{j,n}$ is either equal to $+1$ or -1. If $t_{j,n} = 1$ then $(t_{j,n} - 1)\alpha_n = 0$ and $\gamma_j - \gamma_2 > 0$ by a similar argument as before. If, on the other hand, $t_{j,n} = -1$ and since $j > 2$, then $t_{j,m} = 1$ for some $m < n$ and

$$\gamma_j - \gamma_2 = \sum_{i<n,\, i\neq m} (t_{j,i} + 1)\alpha_i + 2(\alpha_m - \alpha_n)$$

Since the sequence $\{\alpha_i\}$ is arranged in decreasing order we have $(\alpha_m - \alpha_n) > 0$ and $\sum_{i<n,\, i\neq m}(t_{j,i} + 1)\alpha_j \geq 0$ so $\gamma_j - \gamma_2 > 0$.

Similarly, for γ_3 we have $\gamma_3 > \gamma_2 > \gamma_1$, and we need to prove that $\gamma_j - \gamma_3 > 0$ for all $j > 3$. By definition, $\gamma_3 = -\sum_{i\neq n-1} \alpha_i + \alpha_{n-1}$, so for all $j > 3$

$$\gamma_j - \gamma_3 = \sum_{i\neq n-1} (t_{j,i} + 1)\alpha_i + (t_{j,n-1} - 1)\alpha_{n-1}$$

As before, we treat the cases $t_{j,n-1} = 1$ and $t_{j,n-1} = -1$ separately. If $t_{j,n-1} = 1$ then $(t_{j,n-1} - 1)\alpha_{n-1} = 0$ and $\gamma_j - \gamma_i > 0$ since $t_{j,m} > -1$ for some m. On the other hand, $t_{j,n-1} = -1$ and the fact that $j > 3$, implies that there exists at least one $m < n - 1$ such that $t_{j,m} = 1$, so

$$\gamma_j - \gamma_3 = \sum_{i\neq n-1, i\neq m} (t_{j,i} + 1)\alpha_i + 2(\alpha_m - \alpha_{n-1}) > 0.$$

So $\gamma_1 < \gamma_2 < \gamma_3$ are the three smallest elements of the sequence $\{\gamma_i\}$. By the antisymmetric property $\gamma_i = -\gamma_{M-i+1}$, it follows that $\gamma_M > \gamma_{M-1} > \gamma_{M-2}$ are the three largest elements of the same sequence.

Finally, from the definitions of γ_1, γ_2, γ_3, it follows that

$$\begin{aligned} \gamma_2 - \gamma_1 &= 2\alpha_n \\ \gamma_3 - \gamma_1 &= 2\alpha_{n-1} \end{aligned}$$

■

Lemma 3 *Let $n > 2$, and $\boldsymbol{\gamma}$, $\boldsymbol{\alpha}$, M as in Lemma 2. Then*

(a) the sequence γ_j, $j = 1, \cdots, M$, is partitioned in consecutive quadruples where each one is arranged in increasing order, i.e. for all $i = 0, \cdots, 2^{n-2} - 1$, $\gamma_{4i+1} < \gamma_{4i+2} < \gamma_{4i+3} < \gamma_{4i+4}$.

(b) The differences $\gamma_{4i+2} - \gamma_{4i+1} = 2\alpha_n$, $\gamma_{4i+3} - \gamma_{4i+1} = 2\alpha_{n-1}$, $\gamma_{4i+4} - \gamma_{4i+1} = 2(\alpha_{n-1} + \alpha_n)$, are independent of i.

PROOF. For all $i = 0, \cdots, 2^{n-2} - 1$ we have

$$\begin{bmatrix} \gamma_{4i+1} \\ \gamma_{4i+2} \\ \gamma_{4i+3} \\ \gamma_{4i+4} \end{bmatrix} = \begin{bmatrix} b_1 & \cdots & b_{n-2} & -1 & -1 \\ b_1 & \cdots & b_{n-2} & -1 & 1 \\ b_1 & \cdots & b_{n-2} & 1 & -1 \\ b_1 & \cdots & b_{n-2} & 1 & 1 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \vdots \\ \alpha_{n-2} \\ \alpha_{n-1} \\ \alpha_n \end{bmatrix}$$

where $b_1 \cdots b_{n-2}$, corresponds to the binary antipodal representation of i.

Defining $B = [b_1, \cdots, b_{n-2}][\alpha_1, \cdots, \alpha_{n-2}]^T$, we can rewrite the above equation as follows

$$\begin{bmatrix} \gamma_{4i+1} - B \\ \gamma_{4i+2} - B \\ \gamma_{4i+3} - B \\ \gamma_{4i+4} - B \end{bmatrix} = \mathbf{T}^{(2)} \begin{bmatrix} \alpha_{n-1} \\ \alpha_n \end{bmatrix}$$

Then part (a) of the Lemma follows immediately using Lemma 1 while part (b) is easily shown using the structure of $\mathbf{T}^{(2)}$. ■

If the centers c_i are perfectly estimated by $\hat{c}_i$ except for an unknown permutation $\mathbf{\Pi}$ then from Lemmata 1, 2, we have the following corollaries

Corollary 1 *Let $n = 2$ and let the estimated centers $\hat{c}_i$, $i = 1, \cdots, 4$, be arranged in increasing order. Then $\hat{c}_i = c_i$, for all i, and*

$$a_2 = (\hat{c}_2 - \hat{c}_1)/2 \quad (7)$$
$$a_1 = (\hat{c}_3 - \hat{c}_1)/2 \quad (8)$$

Corollary 2 *Let $n = 3$ and let the estimated centers $\hat{c}_i$, $i = 1, \cdots, 8$, be arranged in increasing order. Then $\hat{c}_1 = c_1$, $\hat{c}_2 = c_2$, $\hat{c}_3 = c_3$ and*

$$a_3 = (\hat{c}_2 - \hat{c}_1)/2 \quad (9)$$
$$a_2 = (\hat{c}_3 - \hat{c}_1)/2 \quad (10)$$
$$a_1 = -(\hat{c}_2 + \hat{c}_3)/2 \quad (11)$$

3.3. Recursive Algorithm

Corollaries 1 and 2 tell us how to retrieve the mixing parameters a_i in the cases $n = 2$ or $n = 3$. If there are $n > 3$ sources then the solution can be found using the following recursive algorithm:

Algorithm 1

Step 1. Sort the estimated centers $\hat{c}_i$ in increasing order

Step 2. Estimate a_n, a_{n-1} from

$$\hat{a}_n = (\hat{c}_2 - \hat{c}_1)/2$$

$$\hat{a}_{n-1} = (\hat{c}_3 - \hat{c}_1)/2$$

according to Lemma 2.

Step 3. According to Lemma 3 the sequence $\{c_i\}$ is partitioned in quadruples of the form $[y, y+dc_1, y+dc_2, y+dc_3]$. Estimate the differences dc_i:

$$\hat{dc}_1 = 2\hat{a}_n$$
$$\hat{dc}_2 = 2\hat{a}_{n-1}$$
$$\hat{dc}_3 = 2(\hat{a}_{n-1} + \hat{a}_n)$$

Step 4. Remove the set $\{\hat{c}_1, \hat{c}_2, \hat{c}_3, \hat{c}_1 + \hat{dc}_3\}$ from the sequence $\{\hat{c}_i\}$. If these numbers are not present in the sequence, remove the values that are closest to them. Set $\hat{c}'_1 = \hat{c}_1 + \hat{a}_n + \hat{a}_{n-1}$ as the first element of a new sequence $\{\hat{c}'_i\}$.

Step 5.

Repeat until all elements have been removed:
Find the smallest element $\hat{c}_j$ of the remaining sequence $\{\hat{c}_i\}$;
Remove the set $\{\hat{c}_j, \hat{c}_j + \hat{dc}_1, \hat{c}_j + \hat{dc}_2, \hat{c}_1 + \hat{dc}_3\}$ from $\{\hat{c}_i\}$;
If these numbers are not present in $\{\hat{c}_i\}$, remove the values closest to them;
Keep $\hat{c}_j + \hat{a}_n + \hat{a}_{n-1}$ as the next element of $\{\hat{c}'_i\}$.

At the end, the new sequence $\{\hat{c}'_i\}$ will be 4 times shorter than the original $\hat{c}_i$.

Step 6. Recursively repeat the algorithm for the new sequence $\{\hat{c}'_i\}$ and for a new $n' = n-2$ to obtain $\hat{a}_{n'} = \hat{a}_{n-2}$, $\hat{a}_{n'-1} = \hat{a}_{n-3}$. Eventually we will get $n' = 2$ or $n' = 3$ and the parameters a_1 and a_2, or a_1, a_2, and a_3 can be estimated using Corollaries 1 or 2. ■

To see why the algorithm works, assume that $\hat{c}_i$ are perfect estimates of the true centers c_i except for their ordering. Step 1 is simple while Steps 2, 3 are based on Lemmata 1, 2. Then $\hat{a}_n = a_n$, $\hat{a}_{n-1} = a_{n-1}$, and $\hat{dc}_i = dc_i$, for $i = 1, 2, 3$. Steps 4 and 5 allow the recursive solution of Step 6. Indeed, the minimum element $\hat{c}_j$ of the sequence $\{\hat{c}_i\}$ must be the first element of a quadruple, since $dc_1, dc_2, dc_3 > 0$. Therefore $\hat{c}_j$ has the following form

$$\hat{c}_j = [\pm 1 \cdots \pm 1 \ -1 \ -1][a_1 \cdots a_{n-2} \ a_{n-1} \ a_n]^T$$

By removing one-by-one all the quadruples and keeping the numbers

$$\hat{c}'_i = \hat{c}_j + a_{n-1} + a_n$$

for every minimum element $\hat{c}_j$ of each quadruple we are able to collect numbers of the form $[\pm 1 \cdots \pm 1][a_1 \cdots a_{n-2}]^T$ for all combinations of $+1$ and -1. Therefore, we are able to reduce the problem to a similar problem of smaller dimension, i.e. $\{\hat{c}'_i\}$ corresponds to the vector $\mathbf{T}^{(n-2)}[a_1 \cdots a_{n-2}]^T$. The new dimension n' is less 2 than the dimension n of the original problem. So every recursion in Step 6 will estimate a new pair of mixing parameters: (a_{n-2}, a_{n-3}), (a_{n-4}, a_{n-5}), ... The recursion terminates in the cases where $n' = 3$ or $n' = 2$ where the problem is solved using Corollaries 1 or 2.

3.4. Example

Consider a problem with $n = 4$ sources and mixing vector $\mathbf{a} = [0.88 \ 0.45 \ 0.34 \ 0.27]^T$. The properly ordered true centers are

$$\begin{aligned}\mathbf{c} &= \mathbf{T}^{(4)}\mathbf{a} \\ &= [-1.94, \ -1.40, \ -1.26, \ -0.72, \\ &\quad -1.04, \ -0.50, \ -0.36, \ 0.18, \\ &\quad -0.18, \ 0.36, \ 0.50, \ 1.04, \\ &\quad 0.72, \ 1.26, \ 1.40, \ 1.94]^T \end{aligned} \quad (12)$$

Note, that each quadruple in $\mathbf{c}$ is sorted in increasing order. Suppose now, that the centers are perfectly estimated. Even so, their proper order is lost. Instead, we may produce a sorted list in ascending order (Step 1):

$$\begin{aligned}\hat{\mathbf{c}} &= [-1.94, \ -1.40, \ -1.26, \ -1.04, \\ &\quad -0.72, \ -0.50, \ -0.36, \ -0.18, \\ &\quad 0.18, \ 0.36, \ 0.50, \ 0.72, \\ &\quad 1.04, \ 1.26, \ 1.40, \ 1.94]^T \end{aligned}$$

Then the parameters a_3, a_4, are readily estimated from $\hat{c}_1 = -1.94$, $\hat{c}_2 = -1.40$, $\hat{c}_3 = -1.26$ (Step 2):

$$\hat{a}_4 = (\hat{c}_2 - \hat{c}_1)/2 = 0.27 = a_4$$
$$\hat{a}_3 = (\hat{c}_3 - \hat{c}_1)/2 = 0.34 = a_3$$

Next, we compute the following differences (Step 3):

$$\begin{aligned} dc_1 &= 2\hat{a}_4 = 0.54 \\ dc_2 &= 2\hat{a}_3 = 0.68 \\ dc_3 &= 2(\hat{a}_3 + \hat{a}_4) = 1.22 \end{aligned}$$

It is easy to verify that every quadruple in $\mathbf{c}$ can be written as $[y, y+dc_1, y+dc_2, y+dc_3]$ for some y (ref Eq. (12)).

We remove the set $\{\hat{c}_1, \hat{c}_2, \hat{c}_3, \hat{c}_3 + dc_3\} = \{-1.94, \ -1.40, \ -1.26, \ -1.04\}$ from the sequence $\{\hat{c}\}$ (Step 4) and set the first element of a new sequence $\{\hat{c}'\}$ to be

$$\hat{c}'_1 = \hat{c}_1 + \hat{a}_3 + \hat{a}_4 = -1.33 \ .$$

Repeatedly, we remove elements from $\{\hat{c}\}$ adding new elements to $\{\hat{c}'\}$ as shown below (Step 5):

Iter.	Remove from $\{\hat{c}\}$	Add to $\{\hat{c}'\}$
1	$-1.04, \ -0.50, \ -0.36, \ 0.18$	$\hat{c}'_2 = -0.43$
2	$-0.18, \ 0.36, \ 0.50, \ 1.04$	$\hat{c}'_3 = 0.43$
3	$0.72, \ 1.26, \ 1.40, \ 1.94$	$\hat{c}'_4 = 1.33$

Now, we are faced with a smaller problem with fewer centers $\hat{\mathbf{c}} = [-1.33, \ -0.43, \ 0.43, \ 1.33]^T$ and dimension $n' = n - 2 = 2$. The algorithm is recursively repeated on the new problem (Step 6). In this case the parameters a_1, a_2, can be estimated using Corollary 1:

$$\begin{aligned} \hat{a}_2 &= (\hat{c}'_2 - \hat{c}'_1)/2 = 0.45 = a_2 \\ \hat{a}_1 &= (\hat{c}'_3 - \hat{c}'_1)/2 = 0.88 = a_1 \end{aligned}$$

and the algorithm terminates.

4. RECOVERING THE SOURCES.

Once the mixing vector has been estimated, the sources can be estimated by minimizing

$$\begin{aligned} \hat{\mathbf{s}}(k) &= \arg\min_{\mathbf{s}\in\mathcal{S}} \|x(k) - \hat{\mathbf{a}}^T\mathbf{s}\|^2 \\ &= \arg\min_{\mathbf{t}_i} \|x(k) - \hat{\mathbf{a}}^T\mathbf{t}_i\|^2 \end{aligned} \qquad (13)$$

for every time index k.

5. SIMULATION RESULTS

In this Section we present two sets of experiments: one set without noise and another one with noise.

In the noise-free case we tried different values for n. The centers were estimated with the clustering method outlined in Section 3.1 for $\sigma = 0$. The parameter ϵ was initialized to a small positive value. If the number of centers M obtained were not equal to 2^n the method was repeated using a smaller ϵ, until $M = 2^n$. This approach yields exact estimates for the centers, except that their original order is lost. Table 1 summarizes the results after 10 simulations for each different value of n. The number of samples in each experiment was taken to be $N = 10 \times 2^n$. The performance measures used are the Symbol Error Rate (SER) defined as follows

n	Exec. time (sec)
2	0.04 ± 0.03
4	0.11 ± 0.03
6	0.47 ± 0.03
8	3.73 ± 0.46
10	56.87 ± 5.54

Table 1: Mean and standard deviation of the execution time in seconds after 10 experiments for different numbers of sources, n, in the noise-free case. All experiments gave perfect results results both in terms of SER and the Error Norm.

$$\text{SER} = \frac{\text{Number of erroneous estimated source bits}}{\text{Number of total source bits}} \qquad (14)$$

and the Error Norm, defined as

$$\text{Error Norm} = \|\mathbf{a} - \hat{\mathbf{a}}\| \qquad (15)$$

where the elements of $\mathbf{a}$ are taken to be positive and sorted in descending order.

All experiments in the noise free case showed perfect performance both in terms of SER and in terms of Error Norm. This is expected since the location of the centers is perfectly estimated using the above clustering procedure. Therefore, the mathematical results of section 3.2 hold exactly and the source reconstruction is perfect. Table 1 shows the mean and stadard deviation of the algorithm execution time in a 233MHz Pentium II desktop PC, using MATLAB code. This time includes both the clustering and the actual source separation algorithms. We see that, on the average, the total execution time is less than a minute even for $n = 10$ sources, but of course the time increases exponentially with the number of sources n.

In the second set of experiments we introduce a small amount of noise. It is not our purpose to study the effect of noise on the method performance nor to optimize the clustering procedure for the noisy case. Our aim is rather to show the feasibility of the solution even in the presence of noise using a simple extension of the noise-free clustering algorithm. In particular, we estimate the Gaussian centers by locating the peaks of the histogram of $x(k)$. The advantage of this clustering approach is computational speed. The disadnatage is the fact that it will fail if there are two centers woth distance smaller than σ because then the two Gaussian bells will merge producing a single peak in the histogram. A more advanced approach, e.g. using the EM algorithm, may yield better results, but the study of such clustering methods is beyond the scope of this paper.

Figure 2 shows the result of a noisy experiment with $n = 3$ sources, $N = 1000$ samples, and Signal to Noise Ratio (SNR) = 25.18 db. The performance of the method is SER $= 4 \times 10^{-3}$ and Error Norm $= 0.0149$. The true and estimated mixing vectors are given below

$$\begin{aligned} \mathbf{a} &= [0.651, \ 0.511, \ 0.388]^T \\ \hat{\mathbf{a}} &= [0.648, \ 0.522, \ 0.397]^T \end{aligned}$$

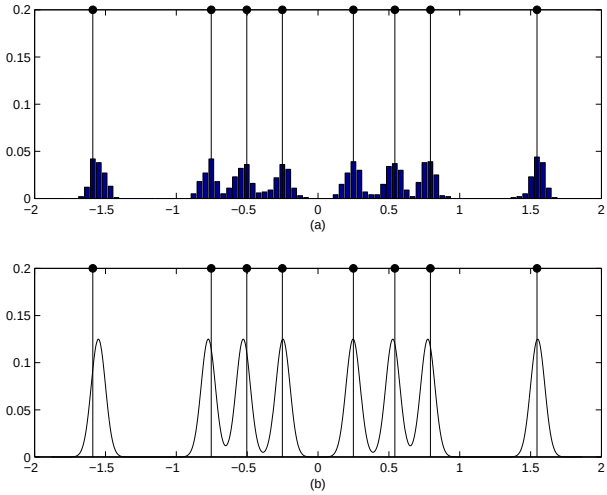

Figure 2: Results of a noisy experiment with $n = 3$ and SNR=25.2db. (a) The histogram of the observed sequence (b) The true distribution of the observed sequence. The vertical lines indicate the estimated center locations.

6. DISCUSSION AND CONCLUSIONS

In this paper we proposed a novel blind method for separating many sources using a single observation sequence. The method applies to binary antipodal sources but it may be extended in the future to other types of sequences with discrete samples, such as mutlilevel PAM or QAM signals. The proposed method differs from standard approaches in that it does not use source statistics, but it is entirely based on the probability distribution of the observation sequence and the fact that it is modelled by a Gaussian mixture.

An issue that needs to be addressed is the estimation of the number of sources n, since this is essential for the accurate performance of Algorithm 1. Fortunately, the number $M = 2^n$ of Gaussian centers in $P(X)$ is related to n, and so an estimate $\hat{M}$ from the histogram of $x(k)$ can provide an estimate $\hat{n}$ as the smallest integer such that $\hat{M} \leq 2^{\hat{n}}$. Such an approach, of course, depends heavily on the performance of the algorithm which detects the number and the locations of the centers from the observations $x(k)$.

Another subtle point is that the method does not work if $\hat{M}$ is not a power of 2. This may happen for two reasons:

(a) there are not enough samples in the observation sequence, or

(b) the estimation of the center locations is poor.

The first problem can be attacked simply by collecting more samples. Clearly, if $N < 2^n$ there will be not enough samples such that every combination vector $\mathbf{t}_i$ will appear at least once in the sequence $\mathbf{s}(1)$, ..., $\mathbf{s}(N)$. As a rule of thumb we need at least $N = 10 \times 2^n$ samples for the algorithm to work well, provided that the combination vectors are equiprobable.

The second problem relates to the effect of noise: as the SNR decreases the performance of the algorithm which estimates the center locations typically deteriorates. For a simple clustering algorithm as the one proposed in Section 5 the problem becomes more pronounced if the distance between two consecutive centers is smaller than σ in which case the respective "bells" merge into a single larger one. This often leads to problems in estimating the hidden centers and to errors. However, more advanced methods (such as the EM algorithm) may lead to better performance against noise. We propose the study of robust center estimation algorithms in the future.

7. REFERENCES

[1] A. Belouchrani, K. Abed-Meraim, J.-F. Cardoso, and E. Moulines, "A Blind Source Separation Technique Using Second-Order Statistics", *IEEE Trans. Signal Processing*, vol. 45, no. 2, pp. 434–444, Feb. 1997.

[2] J.-F. Cardoso, 'Blind Signal Separation: Statistical Principles", *Proc. IEEE*, vol. 86, pp. 2009–2025, Oct. 1998.

[3] P. Comon, "Independent Component Analysis, A New Concept?", *Signal Processing*, vol. 36, no. 3, pp. 287–314, 1994.

[4] A.P. Dempster, N.M. Laird, and D.B. Rubin, "Maximum Likelihood from Incomplete Data via the EM Algorithm", *Journal of the Royal Statistical Society Series B*, vol. 39, pp. 1–38, 1977.

[5] R.O. Duda and P.E. Hart, *Patetrn Classification and Scene Analysis*, John Wiley, NY, 1973.

[6] K.Georgoulakis, S.Theodoridis, "Efficient clustering techniques for channel equalization in hostile environments," *Signal Processing*, vol. 58, pp. 153–164, 1997.

[7] Z. Ghahramani and M. Jordan, "Supervised Learning from Incomplete Data via an EM Approach", in *Advances in Neural Information Processing Systems 6*, J.D. Cowan, G. Tesauro, and J. Alspector (eds.), pp. 120-127, 1994.

[8] S. Haykin, *Neural Networks: A Comprehensive Foundation*, IEEE Press, 1994.

[9] R. Jacobs, M. Jordan, S. Nowlan, and G. Hinton, "Adaptive Mixture of Local Experts", *Neural Computation*, vol. 3, pp. 79–87, 1991.

[10] M. Jordan and R. Jacobs, "Hierarchical Mixtures of Experts and the EM Algorithm", *Neural Computation*, vol. 6, pp. 181–214, 1994.

[11] T.-W. Lee, M.S. Lewicki, M. Girolami, and T.J. Sejnowski, "Blind Source Separation of More Sources than Mixtures Using Overcomplete Representations", *IEEE Singal Processing Letters*, vol. 6, no. 4, April 1999.

[12] D. Lowe, "What Have Neural Network to Offer Statistical Pattern Processing?", in *Proc. SPIE Conf. Adaptive Signal Processing*, pp. 460–471, San Diego, CA, 1991.

[13] G.J. McLachlan and K.E. Basford, *Mixture Models: Inference and Applications to Clustering*, Marcel Dekker, 1988.