Blind Separation of Speech and Sub-Gaussian Signals in Underdetermined Case

SangGyun Kim and Chang D. Yoo

Department of Electrical Engineering and Computer Science
Korea Advanced Institute of Science and Technology, Republic of Korea
zom@eeinfo.kaist.ac.kr and cdyoo@ee.kaist.ac.kr

Abstract

Conventional blind source separation (BSS) algorithms are applicable when the number of sources equals to that of observations; however, they are inapplicable when the number of sources is larger than that of observations. Most underdetermined BSS algorithms have been developed based on an assumption that all sources have sparse distributions. These algorithms are applicable to separate speech signals with super-Gaussian distribution in the underdetermined case. However, they fail to separate the underdetermined mixtures of speech signals and sub-Gaussian signals. In this paper, a novel method for separating the underdetermined mixtures of sources with both super- and sub-Gaussian distributions is proposed. In the proposed method, underdetermined BSS problem is converted to conventional BSS problem by generating hidden observations so that the probability of estimated sources is maximized. Simulation results show that the proposed method can separate the underdetermined mixtures of speech signals and sub-Gaussian signals.

1. Introduction

Independent component analysis (ICA) is a method for finding underlying components that are statistically independent from multivariate statistical data. Recently, blind source separation (BSS) using ICA has received a great deal of attention for its potential in acoustics, telecommunication, and medical and image signal processing. BSS is formulated statistically, that is, data are regarded as random variable $x \in \mathbb{R}^N$ that arises from linear combination of unknown random variables $s \in \mathbb{R}^N$ whose components are mutually independent. It is represented as

$$x = As$$  \hspace{1cm} (1)

where $A$ is an $N \times N$ square matrix. In BSS, the objective is to find an $N \times N$ invertible square matrix $W$ such that

$$u = Wx$$  \hspace{1cm} (2)

where the components of $u$ are as independent as possible. Therefore, the goal of ICA for BSS is to estimate the matrix $A$ and therefore find the independent components (sources) $s$ given only the data (observations) $x$. It is well known that the solution of BSS is allowed up to inherent indeterminations that are permutation and scaling.

Information-maximization (infomax) algorithm based on entropy maximization for BSS has been developed by Bell and Sejnowski [1]. This algorithm is effective in separating sources with super-Gaussian distribution. However, it fails to separate sources with sub-Gaussian distribution. To separate the mixtures of super- and sub-Gaussian sources, Xu et al. [2] and Attias [3] modelled an underlying probability density function of source as a mixture of Gaussians. However, these algorithms are computationally expensive. To simplify the computation and to separate the mixtures of super-Gaussian and sub-Gaussian sources, an extended infomax algorithm was proposed by Lee et al. [4].

In case the number of sources is larger than that of observations, that is, the matrix $A$ is an $M \times N$ matrix with $M < N$, BSS is called as underdetermined BSS. The underdetermined BSS problem is generally more difficult to tackle than the conventional BSS problem where the number of sources is equal to that of observations, since some of the observations are hidden in the underdetermined case. Even when the matrix $A$ is known, the sources $s$ can not be found directly, but have to be inferred. Learning over-complete representations for underdetermined BSS was developed by Lewicki and Sejnowski [5], and it was applied to blind separation of speech signals in the underdetermined case by Lee et al. [6]. However, these methods are based on an assumption that the source is sparsely distributed, in other words, has a super-Gaussian distribution. Therefore, if the assumption is not valid, these methods are not effective. When a source does not satisfy this assumption, a method for achieving the sparsity in a sparser transformed domain, such as by short-time Fourier transform [7] or by wavelet packet transform [8], was proposed; however, these methods do not
guarantee sparsity since achieving sparsity depends on the distribution of the source.

In this paper, a novel method for separating the underdetermined mixtures of sources with both super- and sub-Gaussian distributions is proposed. The proposed method is applicable without the assumption that the sources have sparse distributions. In this method, the underdetermined BSS problem is converted to the conventional BSS problem by generating hidden observations $z$, as shown in Fig. 1. The hidden observation $z$ is generated so that the conditional probability of the hidden observation $z$ given the observation $x$ and the square matrix $W$ is maximized. In Section 3, it will be shown that maximizing the conditional probability of the hidden observations is equivalent to maximizing the summation of the log probabilities of estimated sources. The observations $x$ and the hidden observations $z$ make up a complete data $y$ that is defined as

$$y = \begin{bmatrix} x \\ z \end{bmatrix}$$

where $y \in \mathbb{R}^N$, $x \in \mathbb{R}^M$, and $z \in \mathbb{R}^{N-M}$. With the complete data $y$, conventional BSS algorithms can be applied to estimate the sources in the underdetermined case. Blind separation of speech signals and sub-Gaussian signals is based on extended infomax algorithm proposed by Lee et al. [4]. Laplacian distribution is used to model the distribution of speech signal. A sub-Gaussian distribution is modelled using a symmetrical form of the Pearson mixture model [9]. The hidden observations are generated using these source distributions. The proposed method does not require the assumption that the source distribution is sparse, since the learning of the unmixing matrix $W$ is performed based on a general learning rule of conventional BSS algorithm [1], [4].

This paper is organized as follows. Section 2 reviews conventional underdetermined BSS algorithms using sparse representations. Section 3 shows a proposed method for separating the mixtures of sources that have both sub- and super-Gaussian distributions by generating the hidden data. Section 4 shows simulation results and Section 5 concludes the paper.

### 2. Underdetermined BSS Using Sparse Representations

In the underdetermined BSS model, the sources should be inferred even if the matrix $A$ is known. There are infinitely many solutions to $s$. If the source distribution is sparse, the matrix $A$ can be estimated by either external optimization or clustering and, given $A$, a minimal $l_1$-norm representation of the sources can be obtained by solving a low-dimensional linear programming problem [5], [6], [7], [8]. In these algorithms, even in the case when the matrix $A$ is known, high sparsity is required for good separability. Therefore, these algorithms are not effective in separating the mixtures of the sources, anyone of which has a sub-Gaussian distribution.

### 3. Underdetermined BSS by Generating Hidden Observations

In this paper, the objective is to separate the mixtures of speech signals with super-Gaussian distribution and signals with sub-Gaussian distribution. To achieve this, a novel method for converting the underdetermined BSS problem to the conventional BSS problem by generating hidden observations $z$ is proposed. The hidden observation $z$ is generated by maximizing its conditional probability given the observation $x$ and the matrix $W$. It is represented as following

$$z = \arg \max_z \log p(z|x, W)$$

$$= \arg \max_z \log \frac{p(z|x|W)}{p(x|W)}$$

$$= \arg \max_z \log p(Wy) | \det W|$$

$$= \arg \max_z \sum_i \log p(w_i, y) \quad (4)$$

where $w_i$ is the $i$th row of the matrix $W$ and therefore $w_iy$ represents $u_i$. From (4), the generation of the hidden observations is performed such that the summation of the log-probabilities of the estimated sources is maximized.

After generating the hidden observations $z$, the sources are estimated as a linear product of $W$ and $y$ as in the case of conventional BSS algorithms. It is shown in Fig. 1 and mathematically represented by

$$u = Wy \quad (5)$$

where $u$ are estimated sources.

With the complete data $y$, conventional BSS algorithms can be used in the blind separation of the sources in the underdetermined case. A general learning rule of conventional BSS algorithm, which is given as follows

$$\Delta W \propto [I - \varphi(u)u^T]W \quad (6)$$

where the nonlinearity function $\varphi(u) = \frac{1}{|u|} \frac{\partial \mu(u)}{\partial u}$, can be used.

In order to generate the hidden observations well, the probability density of the source has to be estimated with good precision. In addition, the density estimate of the source plays an important role in the performance of the learning rule of the unmixing matrix. To achieve this, sub- and super-Gaussian density are modelled using different forms. For a sub-Gaussian density, a symmetrical form in [4] is adopted as follows

$$p(u_i) = \frac{1}{2} (N(\mu, \sigma^2) + N(-\mu, \sigma^2)) \quad (7)$$

where $N(\mu, \sigma^2)$ is the normal density with mean $\mu$ and variance $\sigma^2$. In this paper, a sub-Gaussian density with $\mu = 1$
and $\sigma^2 = 1$ is used. For the super-Gaussian density of speech signal, the Laplacian density is used as follows

$$p(u_i) = \frac{1}{\sqrt{2\sigma^2}} e^{-\frac{|u_i|}{\sigma}} \quad (8)$$

where $\sigma^2$ is variance that is set to unity here. The Laplacian density is used since the speech density is described well by the Laplacian density. If the probability density functions of sub- and super-Gaussian sources are determined as (7) and (8), respectively, the nonlinearity function $\varphi(u_i)$ for the general learning rule is given from the density functions as follows

$$\varphi(u_i; k_i) = \begin{cases} \sqrt{2\sigma} \text{sign}(u_i) & \text{for } k_i = 1; \text{ super-Gaussian} \\ u_i - \tanh(u_i) & \text{for } k_i = -1; \text{ sub-Gaussian} \end{cases} \quad (9)$$

where $\text{sign}(u_i)$ gives $1$ when $u_i$ is positive and $-1$ when $u_i$ is negative. Therefore, the nonlinearity function is represented as $\varphi(u_i; k_i)$ where $k_i$ is 1 for super-Gaussian density and -1 for sub-Gaussian density. Switching condition for $k_i$ between the sub- and super-Gaussian distributions is determined according to the sign of kurtosis of estimated source $u_i$; $k_i = 1$ for positive kurtosis and $k_i = -1$ for negative kurtosis.

4. Simulation

In this section, simulation results are shown to verify the performance of the proposed underdetermined BSS algorithm by generating hidden observations in the $2 \times 3$ underdetermined case. Hidden observations are generated to maximize the summation of log probabilities of the estimated sources according to (4). In all simulations, we used a same mixing matrix $A$ which is given as

$$A = \begin{bmatrix} \sqrt{3}/2 & 0 & -1/2 \\ 1/2 & -1 & \sqrt{3}/2 \end{bmatrix} \quad (10)$$

First, separating the $2 \times 3$ underdetermined mixtures of two super-Gaussian sources and one sub-Gaussian source is simulated. The super- and sub-Gaussian sources that are used in the simulation are generated from the hyperbolic-Cauchy density model in [4]; $b = 0$ for super-Gaussian distribution and $b = 2$ for sub-Gaussian, respectively. Some data of length 3000 is used in the learning process that iterates 10 times. The batch size is determined as 42 heuristically. A hidden observation is generated one sample at a time, and the matrix $W$ is updated every batch. Performances comparison between the proposed underdetermined BSS algorithm and the underdetermined algorithm proposed by Lewicki et al. using linear programming (a minimum $\ell_1$-norm solution) [5] is conducted and summarized in Table 1. In Table 1, $\kappa_i^p$, $\kappa_i$, and $\kappa_i^l$ represent the kurtosis of the $i^{th}$ original source, estimated source using the proposed method, and estimated source using linear programming method, respectively; $e_i$ and $e_i^l$ represents the errors using the proposed method and the linear programming after reordering and scaling, respectively, which is given as

$$e_i = \sqrt{\frac{1}{T} \sum_{t=1}^{T} (s_i(t) - u_i(t))^2} \quad (11)$$

where $u_i$ represents the estimate of the $i^{th}$ source $s_i$ and $T$ is the length of the source. The energies of all the sources and their estimates $u$ are normalized to 1 for compensating the scaling effect.

As shown in Table 1, both methods estimate the super-Gaussian sources $s_1$ and $s_2$ to some extend; however, the method proposed by Lewicki et al. fails to separate the sub-Gaussian source $s_3$. The kurtosis of the estimated source $u_3^p$ using the method proposed by Lewicki is positive, and the error $e_3^p$ is much larger than $e_3$ of the proposed method; the Lewicki’s method can not separate the sub-Gaussian source. The kurtosis of the estimated source $u_3$ using the proposed method has a negative value and the error is relatively small; the proposed method can separate the sub-Gaussian sources.

Next, a simulation to separate the $2 \times 3$ underdetermined mixtures of two speech sources with super-Gaussian distribution and one signal with sub-Gaussian distribution is performed. The simulation is conducted with two speech signals and one sub-Gaussian signal sampled at 8 kHz. The length of all the signals is 5000. In Fig. 2, the observations $x$, which are made using the mixing matrix $A$ of (10), and the generated hidden observation $z$ using the proposed method are plotted. The problem of the generation of hidden observation $z$ is solved using a nonlinear optimization subroutine in MATLAB.

In Fig. 3, a simulation result to separate the mixtures of two speech signals and one sub-Gaussian signal in the $2 \times 3$ underdetermined case is shown. As shown in Fig. 3, the proposed algorithm can separate two speech signals and one sub-Gaussian signal from the underdetermined observations.

In general, if a super-Gaussian density is modelled by a unimodal density given in [4], the learning rule of the proposed algorithm can be more simply given as (22) in [4]. However, in this paper, the Laplacian density is used to model super-Gaussian density since it describes the speech density well.

5. Conclusions

A novel method for separating the mixtures of speech signals with super-Gaussian distribution and the signals with

---

Table 1: Performance comparison between the proposed method and linear programming method.

<table>
<thead>
<tr>
<th>Source number</th>
<th>Original Kurtosis</th>
<th>Proposed method</th>
<th>Linear programming</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\kappa_i^p$</td>
<td>$\kappa_i$</td>
<td>$e_i^p$</td>
</tr>
<tr>
<td>1</td>
<td>1.16</td>
<td>0.63</td>
<td>0.52</td>
</tr>
<tr>
<td>2</td>
<td>2.01</td>
<td>0.45</td>
<td>0.48</td>
</tr>
<tr>
<td>3</td>
<td>-1.34</td>
<td>-0.88</td>
<td>0.15</td>
</tr>
</tbody>
</table>

where $e_i^p$ represents the estimate of the $i^{th}$ source $s_i$ and $T$ is the length of the source. The energies of all the sources and their estimates $u$ are normalized to 1 for compensating the scaling effect.

As shown in Table 1, both methods estimate the super-Gaussian sources $s_1$ and $s_2$ to some extent; however, the method proposed by Lewicki et al. fails to separate the sub-Gaussian source $s_3$. The kurtosis of the estimated source $u_3^p$ using the method proposed by Lewicki is positive, and the error $e_3^p$ is much larger than $e_3$ of the proposed method; the Lewicki’s method can not separate the sub-Gaussian source. The kurtosis of the estimated source $u_3$ using the proposed method has a negative value and the error is relatively small; the proposed method can separate the sub-Gaussian sources.

Next, a simulation to separate the $2 \times 3$ underdetermined mixtures of two speech sources with super-Gaussian distribution and one signal with sub-Gaussian distribution is performed. The simulation is conducted with two speech signals and one sub-Gaussian signal sampled at 8 kHz. The length of all the signals is 5000. In Fig. 2, the observations $x$, which are made using the mixing matrix $A$ of (10), and the generated hidden observation $z$ using the proposed method are plotted. The problem of the generation of hidden observation $z$ is solved using a nonlinear optimization subroutine in MATLAB.

In Fig. 3, a simulation result to separate the mixtures of two speech signals and one sub-Gaussian signal in the $2 \times 3$ underdetermined case is shown. As shown in Fig. 3, the proposed algorithm can separate two speech signals and one sub-Gaussian signal from the underdetermined observations.

In general, if a super-Gaussian density is modelled by a unimodal density given in [4], the learning rule of the proposed algorithm can be more simply given as (22) in [4]. However, in this paper, the Laplacian density is used to model super-Gaussian density since it describes the speech density well.
sub-Gaussian distribution is proposed. In this method, the underdetermined BSS problem is converted to conventional BSS problem by generating the hidden observations. The hidden observations are generated so that the probability of estimated source is maximized. In order to separate the speech signals well, the Laplacian density is used to model the speech density. Simulation results show that the proposed method can separate the underdetermined mixtures of speech signals and sub-Gaussian signals.

6. Acknowledgements

This work was supported by grant No. R01-2003-000-10829-0 from the Basic Research Program of the Korea Science and Engineering Foundation and by University IT Research Center Project.

7. References


