Underdetermined Independent Component Analysis by Data Generation

SangGyun Kim and Chang D. Yoo

Department of Electrical Engineering and Computer Science
Korea Advanced Institute of Science and Technology
Guseong-dong, Yuseong-gu, Daejon, Republic of Korea

Abstract. In independent component analysis (ICA), linear transformation that minimizes the dependence among the components is estimated. Conventional ICA algorithms are applicable when the numbers of sources and observations are equal; however, they are inapplicable to the underdetermined case where the number of sources is larger than that of observations. Most underdetermined ICA algorithms have been developed with an assumption that all sources have sparse distributions. In this paper, a novel method for converting the underdetermined ICA problem to the conventional ICA problem is proposed; by generating hidden observation data, the number of the observations can be made to equal that of the sources. The hidden observation data are generated so that the probability of the estimated sources is maximized. The proposed method can be applied to separate the underdetermined mixtures of sources without the assumption that the sources have sparse distribution. Simulation results show that the proposed method separates the underdetermined mixtures of sources with both sub- and super-Gaussian distributions.

1 Introduction

In independent component analysis (ICA), linear transformation to minimize the statistical dependence of the components of the representation is estimated. Recently, blind source separation by ICA has received great deal of attention because of its potential in speech enhancement, telecommunication, and medical signal processing.

In ICA, the objective is to find an $M \times M$ invertible square matrix $\mathbf{W}$ such that

$$\mathbf{s} = \mathbf{W}\mathbf{x} \tag{1}$$

where $\mathbf{s}$ and $\mathbf{x}$ are respectively $M \times 1$ source signal and $M \times 1$ observation, and the components of $\mathbf{s} = \{s_1, s_2, \ldots, s_M\}^T$ are as *independent* as possible. In other words, the j^{th} component x_j of $\mathbf{x}$ can be interpreted as a linear combination of the independent sources since

$$\mathbf{x} = \mathbf{W}^{-1}\mathbf{s} = \mathbf{A}\mathbf{s} \tag{2}$$

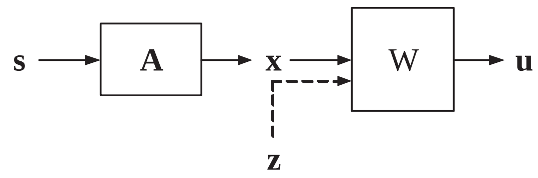

Fig. 1. Block diagram of underdetermined ICA by data generation.

where $\mathbf{A}$ is an $M \times M$ square mixing matrix, and $\mathbf{A} = \mathbf{W}^{-1}$. Therefore, the goal of ICA is to estimate the mixing matrix $\mathbf{A}$ and therefore find the independent sources $\mathbf{s}$ given only the observations $\mathbf{x}$.

Infomax algorithm based on entropy maximization 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 the underlying probability density function (*pdf*) of sources 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].

Conventional ICA algorithms are inapplicable to an underdetermined case where the number of sources is larger than that of observations, that is, the mixing matrix $\mathbf{A}$ is an $m \times M$ matrix with $m < M$. The underdetermined ICA problem is generally more difficult to tackle than the conventional ICA problem where the number of sources is equal to that of observations, since some of the observation data are hidden in the underdetermined case. Even if the mixing matrix $\mathbf{A}$ is estimated exactly, the sources $\mathbf{s}$ can not be found directly, but have to be inferred [5]. The overcomplete representation and sparse coding were studied by Olshausen and Field [6] and were later developed as learning overcomplete representations for ICA by Lewicki and Sejnowski [5]. This method was applied to blind separation of speech signals in the underdetermined case by Lee et al. [7]. However, methods such as these are based on an assumption that the distribution of the source is sparse. 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 [8] or by wavelet packet transform [9], was proposed. However, the method does not guarantee sparsity since achieving sparsity depends on the distributions of the sources.

In this paper, a novel method for converting the underdetermined ICA problem to the conventional ICA problem by generating the hidden observation data $\mathbf{z}$, as shown in Fig. 1, is proposed. The hidden data $\mathbf{z}$ is generated so that the conditional probability of the hidden data $\mathbf{z}$ given the observation $\mathbf{x}$ and the unmixing matrix $\mathbf{W}$ is maximized. The observation data $\mathbf{x}$ and the hidden data $\mathbf{z}$ make up a complete data $\mathbf{y}$ that is defined as

$$\mathbf{y} \equiv \begin{bmatrix} \mathbf{x} \\ \mathbf{z} \end{bmatrix} \tag{3}$$

where $\mathbf{y} \in \mathbb{R}^M$ and $\mathbf{z} \in \mathbb{R}^{M-m}$. With the complete data $\mathbf{y}$, conventional ICA algorithms can be applied to estimate the sources in the underdetermined case. In order to separate the mixtures of sources that have both sub- and super-Gaussian distribution, the hyperbolic-Cauchy density model in [4], which can describe both super- and sub-Gaussian distribution, is used to model the *pdf* of source. The proposed method does not require the assumption that the source distribution is sparse since the learning of the square unmixing matrix $\mathbf{W}$ is performed based on the extended infomax ICA algorithm [4].

This paper is organized as follows. Section 2 presents the extended infomax algorithm proposed by Lee et al. [4]. Section 3 presents the proposed underdetermined ICA algorithm by data generation. Section 4 shows the simulation results. Section 5 discusses the problem of the proposed method, and Section 6 concludes the paper.

2 The Extended Infomax Algorithm

An unsupervised learning algorithm based on entropy maximization was proposed by Bell and Sejnowski [1]. This algorithm is effective in separating sources that have super-Gaussian distribution. However, it fails to separate sources that have sub-Gaussian distribution. In order to separate the mixtures of super-Gaussian and sub-Gaussian sources, an extended infomax algorithm is proposed preserving the simple architecture of infomax algorithm by Lee et al. [4]. It provides a simple learning rule with a parametric density model that can have various distributions by changing the value of a parameter. One proposed parametric density that may be used to model both sub- and super-Gaussian source data s is given as

$$p_s(s) = \frac{1}{4}\{sech^2(s+b) + sech^2(s-b)\} \tag{4}$$

where b is a constant. Depending on the value that b takes, $p_s(s)$ can model either sub- or super-Gaussian distribution. For example, when b=0, the parametric density is proportional to the hyperbolic-Cauchy distribution and therefore is suited for separating super-Gaussian distributions. When b=2, it has a bimodal distribution with negative kurtosis and therefore is suited for separating sub-Gaussian distributions. Switching between the sub- and super-Gaussian is determined according to the sufficient condition that guarantees asymptotic stability [10].

3 Underdetermined ICA by Data Generation

When the number of sources is larger than that of observations, it is difficult to estimate the sources given only the observations. In this section, sparse representation for the underdetermined ICA is briefly reviewed [5], [7], [8], [9], and a novel algorithm for underdetermined ICA when the sources have the super- and sub-Gaussian distributions is proposed.

3.1 Underdetermined ICA Using Sparse Representations

In the underdetermined ICA model, the sources should be inferred even if the mixing matrix, $\mathbf{A}$, is known. There are infinitely many solutions to $\mathbf{s}$. If the source distribution is sparse, the mixing matrix can be estimated by either external optimization or clustering and, given the mixing matrix, a minimal l_1-norm representation of the sources can be obtained by solving a low-dimensional linear programming problem [5], [7], [8], [9]. In these algorithms, even in the case when the mixing matrix 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.

3.2 Underdetermined ICA by Data Generation

In this paper, the objective is to separate the underdetermined mixtures of sources that have sub- and super-Gaussian distributions. To achieve this, a novel method for converting the underdetermined ICA problem to the conventional ICA problem by generating hidden observations $\mathbf{z}$ is proposed. The hidden data $\mathbf{z}$ is generated by maximizing the conditional probability of the hidden data $\mathbf{z}$ given the observation, $\mathbf{x}$, and the unmixing matrix, $\mathbf{W}$. It is given as following

$$\begin{aligned}
\mathbf{z} &= arg \max_{\mathbf{z}} \ \log p(\mathbf{z}|\mathbf{x},\mathbf{W}) && (5)\\
&= arg \max_{\mathbf{z}} \ \log \frac{p(\mathbf{z},\mathbf{x}|\mathbf{W})}{p(\mathbf{x}|\mathbf{W})} && (6)\\
&= arg \max_{\mathbf{z}} \ \log p_{\mathbf{s}}\left(\mathbf{W}\mathbf{y}\right)|\det \mathbf{W}| && (7)\\
&= arg \max_{\mathbf{z}} \ \sum_{i}^{M} \log p_{s_i}\left(\mathbf{w}_i \mathbf{y}\right) && (8)
\end{aligned}$$

where $\mathbf{w}_i$ is the i^{th} row of the unmixing matrix $\mathbf{W}$ and $\mathbf{y}$ is given in (3). From (8), the generation of the hidden data is performed such that the summation of the log-probabilities of the estimated sources is maximized.

After generating the hidden data $\mathbf{z}$, as shown in Fig. 1, the sources are estimated as a linear product of $\mathbf{W}$ and $\mathbf{y}$ defined in (3) as in the case of conventional ICA algorithms. This is mathematically represented by

$$\mathbf{u} = \mathbf{W}\mathbf{y} \qquad (9)$$

where $\mathbf{W}$ is an $M \times M$ unmixing matrix and $\mathbf{u}$ are estimated sources.

In order to generate the hidden data well, the probability density of the sources 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 mixing matrix. To achieve this, the parametric density of (4) is used to model the source distribution. The $M \times 1$ parameter $\mathbf{b}$ representing column vector of constant b in (4) should be updated so as to match the parametric density function to the source density function. Therefore, the learning rule for $\mathbf{b}$ is given as

$$\Delta \mathbf{b} \propto \frac{\partial \log p_s(\mathbf{u}|\mathbf{b})}{\partial \mathbf{b}} \tag{10}$$

to maximize the log-likelihood of the source.

Next, the learning algorithm for the unmixing matrix $\mathbf{W}$ for sub- and super-Gaussian sources is

$$\Delta \mathbf{W} \propto [\mathbf{I} + 2\tanh(\mathbf{u})\mathbf{u}^T - 2\tanh(\mathbf{u}+\mathbf{b})\mathbf{u}^T - 2\tanh(\mathbf{u}-\mathbf{b})\mathbf{u}^T]\mathbf{W} \tag{11}$$

which was given in [4].

Therefore, the underdetermined ICA algorithm by data generation is summarized as follows. First, the unmixing matrix $\mathbf{W}$ and the parameter $\mathbf{b}$ of the source density is initialized, respectively. After initialization, the hidden data $\mathbf{z}$ is generated to maximize the summation of the log-probabilities of the estimated sources according to (8) given the observations $\mathbf{x}$ and the unmixing matrix $\mathbf{W}$. After generating the hidden data, the source $\mathbf{s}$ is estimated according to (9) and then the parameter $\mathbf{b}$ and the unmixing matrix $\mathbf{W}$ is updated according to (10) and (11), respectively. Finally, at next iteration, we start again from the data generation using $\mathbf{W}$ and $\mathbf{b}$ of the previous step.

4 Simulation Results

In this section, simulation results are shown to verify the performance of the underdetermined ICA algorithm by data generation for the 2×3 underdetermined case. In Example 1, the performances of two algorithms to separate the underdetermined mixtures of 2 sources of super-Gaussian distributions and 1 source of sub-Gaussian distribution are compared. One algorithm is the proposed underdetermined ICA algorithm by data generation and the other is the underdetermined ICA algorithm based on the minimum l_1-norm solution using the linear programming [5]. In Example 2, it is shown that the proposed algorithm can separate the mixtures of two speech signals and one sub-Gaussian signal.

In all experiments, a same mixing matrix $\mathbf{A}$ is used, which is given as

$$\mathbf{A} = \begin{bmatrix} 1/\sqrt(2) & -1/\sqrt(2) & 1 \\ 1/\sqrt(2) & 1/\sqrt(2) & 0 \end{bmatrix}. \tag{12}$$

The problem of generating the hidden data is solved using a nonlinear optimization subroutine in MATLAB. The hidden data generation based on (8) is

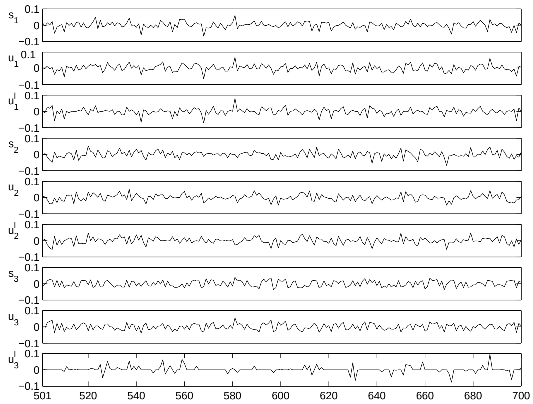

Fig. 2. Separation of the mixtures of the sources that have super- and sub-Gaussian distributions using the proposed and Lewicki's algorithm.

actually performed so that $-\sum_i^M \log p_{s_i}(\mathbf{w}_i\mathbf{y})$ is minimized using the nonlinear optimization (minimization) function in MATLAB.

Example 1: The simulation of separating the 2×3 underdetermined mixtures of the sources that have different distributions is performed. Two sources s_1 and s_2 have the super-Gaussian distributions, and the other s_3 has the sub-Gaussian distribution. The super- and sub-Gaussian sources that are used in the simulation are generated from the hyperbolic-Cauchy density model of (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 100. A batch hidden data are generated one sample at a time using same unmixing matrix for that batch. The unmixing matrix and the density parameter are updated every batch. The learning rates for the unmixing matrix and the density parameter are 0.001 and 0.001, respectively.

In Fig. 2, the simulation result using the proposed and Lewicki's algorithm is shown after reordering and rescaling. In Fig. 2, s_i, u_i, and u_i^l represent the i^{th} original source, the estimate of s_i using the proposed algorithm, and the estimate of s_i using the algorithm proposed by Lewicki et al. for i =1, 2, and 3, respectively. The sources that have super-Gaussian distributions are estimated to some extent in both algorithms, however, the source of sub-Gaussian distribution is estimated well only when using the proposed algorithm as expected.

In Table 1, the simulation results are summarized; κ_i^o, κ_i, and κ_i^l represents the kurtosis of the i^{th} original source, estimated source using the proposed method, and estimated source using Lewicki's method, respectively, and $corr_i$ and $corr_i^l$ represent the correlation coefficient between the original source signal and the estimated source signal using the proposed method and Lewicki's method after reordering, respectively. As shown in Table 1, it is also verified that both methods can estimated the super-Gaussian sources; however, the method proposed by Lewicki et al. fails to separate the sub-Gaussian source. The kurto-

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

Source number	Original kurtosis	Proposed method		Lewicki's method	
	κ_i^o	κ_i	$corr_i$	κ_i^l	$corr_i^l$
1	1.16	0.63	0.74	1.65	0.66
2	2.01	0.45	0.76	1.04	0.78
3	-1.34	-0.88	0.92	9.30	0.04

sis κ_3^l of the estimated source u_3^l is positive, and the correlation coefficient $corr_3^l$ between the sub-Gaussian signal s_3 and the estimated signal u_3^l is very small.

Example 2: Finally, the proposed algorithm is applied to separate the mixtures of two speech signals and one noise that have sub-Gaussian distribution. Fig. 3 shows the separation results. It is shown that the proposed algorithm separates two speech signals and the sub-Gaussian noise to some extent as in Example 2.

5 Discussion

The proposed method is based on a parametric density model. Therefore, when the parametric density model of (4) does not describe the source densities well, e.g., speech density, the performance of the proposed method is degraded. That is also verified in Example 2. In order to solve this problem, nonparametric density estimation method in [11] can be applied to this method. In that case, the generation equation of (8) and the learning rule of (11) should be modified based on the estimated nonparametric density. Further study is need to obtain an underdetermined ICA algorithm using nonparametric density estimation.

6 Conclusion

A novel method for applying the extended infomax algorithm to the underdetermined ICA model is proposed. This is achieved by converting the underdetermined ICA problem to the conventional ICA problem by generating the hidden observation data. The hidden data are generated to maximize the summation of the log-probabilities of the estimated sources. The simulation results show that the proposed algorithm can separate the underdetermined mixtures of the sources that have sub- and super-Gaussian distributions. However, further study is needed to determine until what dimensionality of the hidden data the proposed algorithm is effective and modify this algorithm to be nonparametric.

7 Acknowledgement

This work was supported in part 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.

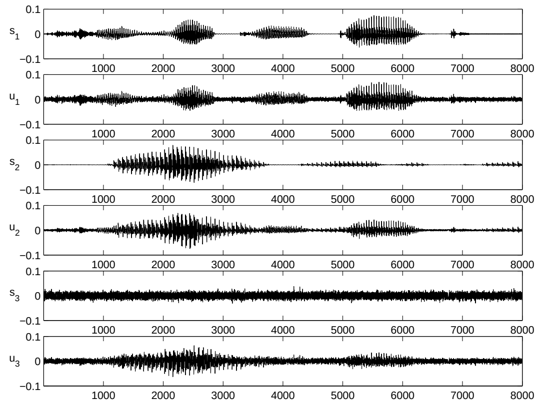

Fig. 3. Separation of the mixtures of two speech signals with super-Gaussian distribution and one noise with sub-Gaussian distribution.

References

1. Bell, A.J. and Sejnowski, T.J.: An information-maximisation approach to blind separation and blind deconvolution. Neural Computation, Vol. 7, No. 6. (1995) 1129–1159
2. Xu, L., Cheung, C., Yang, H., and Amari, S.: Maximum equalization by entropy maximization and mixture of cumulative distribution functions. Proceedings of ICNN. Houston. (1997) 1821–1826
3. Attias, H.: Independent factor analysis. Neural Computation. Vol. 11. (1999) 803–852
4. Lee, T.W., Girolami, M., and Sejnowski, T.J.: Independent component analysis using an extended infomax algorithm for mixed sub-Gausian and super-Gaussian sources. Neural Computation. Vol. 11, No. 2. (1999) 409–433
5. Lewicki, M.S. and Sejnowski, T.J.: Learning overcomplete representations. Neural Computation. Vol. 12. (2000) 337–365
6. Olshausen, B.A. and Field D.J.: Sparse coding with an overcomplete basis set: A strategy employed by V1?. Vision Research. Vol.11. (1997) 3311–3325
7. Lee, T.W., Lewicki, M.S., Girolami, M., and Sejnowski, T.J.: Blind source separation of more sources than mixtures using overcomplete representation. IEEE Signal Processing Letters. Vol. 6, No. 4. (1999) 87–90
8. Bofill, P. and Zibulevsky, M.: Underdertermined blind source separation using sparse representations. Signal Processing. Vol. 81. (2001) 2353–2362
9. Li, Y., Cichocki, A., and Amari, S.I.: Sparse component analysis for blind source separation with less sensors than sources. 4th International Symposium on Independent Component Analysis and Blind Signal Separation. Nara Japan (2003) 89–94
10. Cardoso, J.F.: Blind signal processing: statistical principles. Proceedings of the IEEE. (1998) 2009–2025
11. Boscolo, R., Pan, H., and Roychowdhury, V. P.: Independent component analysis based on nonparametric density estimation. IEEE transaction on neural networks, Vol. 15, No. 1. (2004) 55–65