Underdetermined Blind Source Separation Based on Subspace Representation

Citation: SangGyun Kim, and C.D. Yoo. "Underdetermined Blind Source Separation Based on Subspace Representation." Signal Processing, IEEE Transactions on 57.7 (2009): 2604-2614.

As Published: http://dx.doi.org/10.1109/tsp.2009.2017570

Publisher: Institute of Electrical and Electronics Engineers

Persistent URL: http://hdl.handle.net/1721.1/51862

Version: Final published version: final published article, as it appeared in a journal, conference proceedings, or other formally published context

Underdetermined Blind Source Separation Based on Subspace Representation

SangGyun Kim, *Member, IEEE*, and Chang D. Yoo, *Member, IEEE*

***Abstract*—This paper considers the problem of blindly separating sub- and super-Gaussian sources from underdetermined mixtures. The underlying sources are assumed to be composed of two orthogonal components: one lying in the rowspace and the other in the nullspace of a mixing matrix. The mapping from the rowspace component to the mixtures by the mixing matrix is invertible using the pseudo-inverse of the mixing matrix. The mapping from the nullspace component to zero by the mixing matrix is noninvertible, and there are infinitely many solutions to the nullspace component. The latent nullspace component, which is of lower complexity than the underlying sources, is estimated based on a mean square error (MSE) criterion. This leads to a source estimator that is optimal in the MSE sense. In order to characterize and model sub- and super-Gaussian source distributions, the parametric generalized Gaussian distribution is used. The distribution parameters are estimated based on the expectation-maximization (EM) algorithm. When the mixing matrix is unavailable, it must be estimated, and a novel algorithm based on a single source detection algorithm, which detects time-frequency regions of single-source-occupancy, is proposed. In our simulations, the proposed algorithm, compared to other conventional algorithms, estimated the mixing matrix with higher accuracy and separated various sources with higher signal-to-interference ratio.**

***Index Terms*—Generalized Gaussian distribution, single source detection, subspace representation, underdetermined blind source separation (BSS).**

I. Introduction

A blind source separation (BSS) aims to recover unobserved sources from a number of observed mixtures without knowing the mixing system, and it is generally performed with some *a priori* knowledge of the mixing system. It has received considerable attention for its potential applications in speech processing, biomedical processing, and digital communications [1], [2]. Today, there exist a variety of BSS algorithms, and most exploit one of the following four properties of the sources: the higher-order statistics (HOS), the second-order statistics (SOS), the nonstationarity, and the sparsity [3]. More specifically, the algorithms can exploit one of the following properties: mutual independence between the sources [4]–[6], the temporal structure of the sources [7], [8], the temporal structure of the sources [9], [10], and the sparse representation of the sources [11]–[19].

When the number of the sources is larger than that of the mixtures, the BSS problem is called an underdetermined BSS problem. This problem is generally more difficult than the complete BSS problem where the number of the sources is equal to that of the mixtures. In the underdetermined case, the sources are not obtained easily and must be inferred even when the mixing matrix is known.

Of the four properties exploited, most conventional underdetermined BSS algorithms were developed based on the sparse representation of the sources [11]–[19]. Chen *et al.* considered sparse representation of signals given overcomplete dictionaries [11]. Donoho *et al.* considered the sparse representation of the sources via l_1-norm minimization [12]. Lewicki *et al.* developed an algorithm for learning overcomplete representation of the sources based on the maximum *a posteriori* (MAP) estimation of the sparse sources [13]. Lee *et al.* applied this algorithm on speech [14]. Zibulevsky *et al.* also estimated the mixing matrix and sources based on both the MAP and the maximum likelihood (ML) estimation [15]. Girolami proposed a variational expectation-maximization (EM) algorithm for learning sparse and overcomplete representations [16]. Recently, algorithms for achieving the sparsity in transform domain, such as by wavelet packet transform [17] or by short-time Fourier transform (STFT) [18], [19], were proposed. Yilmaz *et al.* assumed that the sources are disjoint in the time-frequency (TF) domain, i.e., there exists only one source at any TF point [18]. Aïssa-El-Bey *et al.* relaxed this condition and assumed that the sources can be nondisjoint in the TF domain, but the number of the sources that coexist at any TF point is less than that of the mixtures [19]. Both algorithms work well on speech.

Recently, a number of Bayesian based algorithms using flexible distributions were proposed [20]–[22]. To model the underlying source distributions, Cemgil *et al.* used the student t-distribution [20], [21], and Snoussi *et al.* used the generalized hyperbolic distribution [22], respectively. The student t-distribution can model the distribution of a sparse source but can not model the distribution of a nonsparse source. The generalized hyperbolic distribution can model a wider range of distributions but requires the estimation of five parameters. These algorithms consider background noise and a wide variety of source distributions in their formulation; however, they require high computational load since they simultaneously estimate the underlying

Manuscript received August 27, 2008; revised February 04, 2009. First published March 16, 2009; current version published June 17, 2009. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Danilo P. Mandic. This work was supported by Grant R01-2007-000-20949-0 from the Basic Research Program of the Korea Science and Engineering Foundation and by Brain Korea 21 Project, the School of Information Technology, KAIST in 2008

S. Kim is with the Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139-4307 USA (e-mail: sgkim1@mit.edu).

C. D. Yoo is with the Division of Electrical Engineering, School of Electrical Engineering and Computer Science, Korea Advanced Institute of Science and Technology, Daejeon 305-701, Korea (e-mail: cdyoo@ee.kaist.ac.kr).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSP.2009.2017570

sources as well as all the parameters involved in the sampling. Another concern with these algorithms is that very slow convergence rate is observed with a small background noise: the covariance of the conditional distribution of the parameter is inversely proportional to noise variance, and the Markov chain often used in sampling can not efficiently explore the parameter domain [22].

This paper considers the problem of blindly separating sources of sub- and super-Gaussian distributions from underdetermined mixtures. The proposed algorithm is derived without considering a background noise as in [20]–[22]. However, the robustness of the algorithm in background noise is considered. In instantaneous mixing, the sources can be represented as the sum of two orthogonal components: one lies in the rowspace and the other in the nullspace of the mixing matrix. The mapping from the rowspace component to the mixtures by the mixing matrix is invertible using the pseudo-inverse of the mixing matrix. The mapping from the nullspace component to zero by the mixing matrix is noninvertible, and there are infinitely many solutions to the nullspace component. The nullspace component is latent. The proposed algorithm estimates the latent nullspace component to minimize the mean-square error (MSE) between its true and estimated value using a certain source distribution. This paper shows that the latent nullspace component estimation (LNCE) based on the minimum MSE (MMSE) leads to overall source estimation that satisfies the MSE criterion.

The proposed underdetermined BSS algorithm is based on a three-stage approach whose overall structure is shown in Fig. 1. Firstly, a mixing matrix is estimated given only the mixtures. A novel single source detection (SSD) algorithm, which detects TF points occupied by only a single source for each source, is proposed. The mixing matrix is estimated using the mixtures in the detected TF points. The proposed algorithm can estimate the mixing matrix with only a single TF point of single-source-occupancy (SSO) for each source. Secondly, the underlying source distributions are estimated using the EM algorithm. In order to model the underlying sources, various parametric distributions such as the hyperbolic-Cauchy distribution (HCD) [23] and the generalized Gaussian distribution (GGD) [24] were considered. For many signals encountered in BSS such as speech, audio, EEG, etc., we have found that the GGD can model the signal distribution better than the HCD that is inadequate for modeling either highly sparse signals or uni-modal nonsparse signals. For this reason, we have chosen the GGD to model the signal distribution. The parameters of the GGD are estimated based on the EM algorithm by considering the nullspace component of the sources as a latent variable. Finally, the underlying sources are estimated given the mixtures, the estimated mixing matrix, and the estimated source distributions. In [23], the underdetermined BSS problem is transformed to a complete BSS problem by generating latent mixture to maximize its conditional probability without regard to any optimality condition. However, this paper transforms the underdetermined BSS problem into the LNCE problem and estimates the nullspace component to minimize the MSE of the sources. This transformation also reduces a computational load since only the nullspace component of the sources is estimated instead of the sources directly. For example, in the 3×4 underdetermined case we should estimate four sources, but by this transformation we need to estimate only one nullspace component.

The rest of this paper is organized as follows. Section II formulates the problem that we are addressing. Section III describes how the sources are represented using the fundamental subspaces of the mixing matrix. Section IV proposes an algorithm for estimating the mixing matrix based on the SSD in the TF domain. Section V describes the GGD and presents an algorithm for updating the parameters of the GGD based on the EM algorithm. Section VI estimates the sources by the LNCE based on the MMSE. Section VII shows the simulation results, and Section VIII concludes the paper.

II. Problem Formulation

The general goal of the BSS is to find an unmixing system to recover the unobserved sources from the observed mixtures without knowing the mixing system. There are various ways of formulating a BSS problem depending on how the sources are mixed, how many sources and mixtures there are, whether there exists a noise or not, etc. Depending on the problem formulation, a different approach is required. The problem considered in this paper is an underdetermined instantaneous BSS without a background noise and can be mathematically formulated as follows.

Let $\boldsymbol{x}[n] = \begin{bmatrix} x_1[n] & \cdots & x_m[n] \end{bmatrix}^T$ be an m-dimensional random variable of the mixtures and $\boldsymbol{s}[n] = \begin{bmatrix} s_1[n] & \cdots & s_M[n] \end{bmatrix}^T$ be an M-dimensional random variable of the nth sample of the source signals. In the absence of noise, the relationship between $\boldsymbol{x}[n]$ and $\boldsymbol{s}[n]$ is given by

$$\boldsymbol{x}[n] = \mathbf{A}\boldsymbol{s}[n] \tag{1}$$

where $\mathbf{A}$ is an $m \times M$ mixing matrix. When the number of the mixtures is less than that of the sources ($m < M$), the BSS is categorized as underdetermined. In such case, the sources can not be obtained directly as in the complete BSS case, since the inverse of $\mathbf{A}$ does not exist.

This paper represents the underlying sources as two orthogonal components: one lies in the rowspace and the other in the nullspace of $\mathbf{A}$ and investigates how the latent nullspace component is estimated using the underlying source distribution. In addition, novel algorithms for estimating $\mathbf{A}$ based on the SSD algorithm and estimating the parameters of the source distribution based on the EM algorithm are introduced.

III. Representation of Sources Using Fundamental Subspaces of $\mathbf{A}$

This section explains how the solution of the underdetermined BSS problem can be represented using $\boldsymbol{x}$ when $\mathbf{A}$ is known. This representation can also be explained graphically using the fundamental subspaces of $\mathbf{A}$.

A general solution of underdetermined BSS problem can be given as the sum of a particular solution of the nonhomogeneous equation $\mathbf{A}\boldsymbol{s} = \boldsymbol{x}$ and a general solution of the homogeneous equation $\mathbf{A}\boldsymbol{s} = \mathbf{0}$ [25]. It can be expressed in the following form:

$$\boldsymbol{s} = \mathbf{A}^{-}\boldsymbol{x} + \mathbf{V}\boldsymbol{z} \tag{2}$$

where $\mathbf{A}^-$ denotes the generalized-inverse (g-inverse) of $\mathbf{A}$ such that $\hat{\boldsymbol{s}} = \mathbf{A}^-\boldsymbol{x}$ is a solution of $\boldsymbol{s}$ of $\mathbf{A}\boldsymbol{s} = \boldsymbol{x}$ for any $\boldsymbol{x}$ which makes this equation consistent, $\mathbf{V}$ is an $M \times (M-m)$ matrix whose columns are bases of the nullspace of $\mathbf{A}$, and $\boldsymbol{z}$ is an $(M-m) \times 1$ arbitrary vector, respectively. The basis matrix $\mathbf{V}$ for the nullspace of $\mathbf{A}$ can be obtained from $\mathbf{A}$. That is, the general solution of $\mathbf{A}\boldsymbol{s} = \boldsymbol{x}$ can be represented as the sum of the particular nonhomogeneous solution $\mathbf{A}^-\boldsymbol{x}$ and the general homogeneous solution $\mathbf{V}\boldsymbol{z}$. Donoho *et al.* also represented a solution of underdetermined linear equation using the nullspace of $\mathbf{A}$ and estimated the solution that minimized the Kolmogorov-complexity, leading to a sparse solution [26].

In the underdetermined BSS problem, there are infinitely many solutions to $\boldsymbol{s}$, since $\mathbf{A}^-$ and $\boldsymbol{z}$ are not unique ($\boldsymbol{z}$ is an arbitrary vector in $\mathbb{R}^{M-m}$). The Moore-Penrose pseudo-inverse matrix that leads to the least-squares solution of $\boldsymbol{s}$ is often used as $\mathbf{A}^-$. When $rank(\mathbf{A}) = m$, the Moore–Penrose pseudo-inverse matrix, denoted as $\mathbf{A}^\dagger$, is given as follows:

$$\mathbf{A}^\dagger = \mathbf{A}^T(\mathbf{A}\mathbf{A}^T)^{-1}. \tag{3}$$

Using $\mathbf{A}^\dagger$, the general solution is given as

$$\boldsymbol{s} = \mathbf{A}^\dagger\boldsymbol{x} + \mathbf{V}\boldsymbol{z}. \tag{4}$$

As represented in (4), the problem of estimating $\boldsymbol{s}$ boils down to the problem of estimating $\boldsymbol{z}$, which leads to the reduction of the dimension of an estimator from M of $\hat{\mathbf{s}}$ to $M-m$ of $\hat{z}$. The representation of (4) can also be explained graphically using the four fundamental subspaces of $\mathbf{A}$: *columnspace*, *nullspace*, *rowspace*, and *left nullspace* of $\mathbf{A}$. Each space is defined in [27]. The four fundamental subspaces of $\mathbf{A}$ have the following important property.

1) Property 1: $\mathbb{R}^M$ can be decomposed into the rowspace and the nullspace of $\mathbf{A}$, and $\mathbb{R}^m$ can be decomposed into the columnspace and the left nullspace of $\mathbf{A}$, respectively. This property is proven in [28].

By *Property 1*, a source vector $\boldsymbol{s} \in \mathbb{R}^M$ can be decomposed into its rowspace component $\boldsymbol{s}_r$ and its nullspace component $\mathbf{s}_n$ as follows:

$$\boldsymbol{s} = \boldsymbol{s}_r + \mathbf{s}_n \tag{5}$$

for a given $\mathbf{A}$. Comparing (4) to (5), $\mathbf{A}^\dagger\boldsymbol{x}$ and $\mathbf{V}\mathbf{z}$ in the right-hand side of (4) correspond to $\boldsymbol{s}_r$ and $\boldsymbol{s}_n$ in the right-hand side of (5), respectively. This is proven as follows: First, $\mathbf{A}^\dagger\boldsymbol{x}$ on the right-hand side of (4) is represented as $\mathbf{A}^\dagger\boldsymbol{x} = \mathbf{A}^T[(\mathbf{A}\mathbf{A}^T)^{-1}\boldsymbol{x}] = \mathbf{A}^T\boldsymbol{y}$ where $\boldsymbol{y} = (\mathbf{A}\mathbf{A}^T)^{-1}\boldsymbol{x}$. Since $\boldsymbol{y} \in \mathbb{R}^m$, $\mathbf{A}^\dagger\boldsymbol{x}$ is in the rowspace of $\mathbf{A}$ [27]. Second, since $\mathbf{A}\mathbf{V}\boldsymbol{z} = \mathbf{0}\boldsymbol{z} = \mathbf{0}$, $\mathbf{V}\boldsymbol{z}$ is in the nullspace of $\mathbf{A}$ [27].

The relationships among the fundamental subspaces of $\mathbf{A}$ is summarized in Fig. 2. On the left-hand side is the space $\mathbb{R}^M$, which can be decomposed into the rowspace and the nullspace of $\mathbf{A}$, which are orthogonal. The transformation on $\boldsymbol{s}$ by $\mathbf{A}$ represented as $\mathbf{A}\boldsymbol{s} = \mathbf{A}\boldsymbol{s}_r + \mathbf{A}\boldsymbol{s}_n$ is explained in the two steps. First, the rowspace component is mapped to the columnspace: $\mathbf{A}\boldsymbol{s}_r = \mathbf{A}\boldsymbol{s}$ is in the columnspace of $\mathbf{A}$. Second, the nullspace component is mapped to zero: $\mathbf{A}\boldsymbol{s}_n = \mathbf{0}$. On the right-hand side of Fig. 2 is the space $\mathbb{R}^m$, which is decomposed into the orthogonal spaces of the columnspace and the left nullspace of $\mathbf{A}$. The mixing matrix $\mathbf{A}$ maps its rowspace to its columnspace. The mapping from the rowspace to the columnspace is invertible using $\mathbf{A}^\dagger$. However, the mapping from $\mathbf{0}$ to the nullspace is noninvertible.

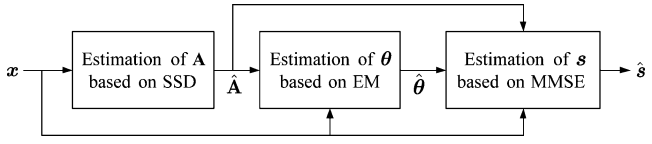


Fig. 1. Block diagram of the proposed underdetermined BSS algorithm.

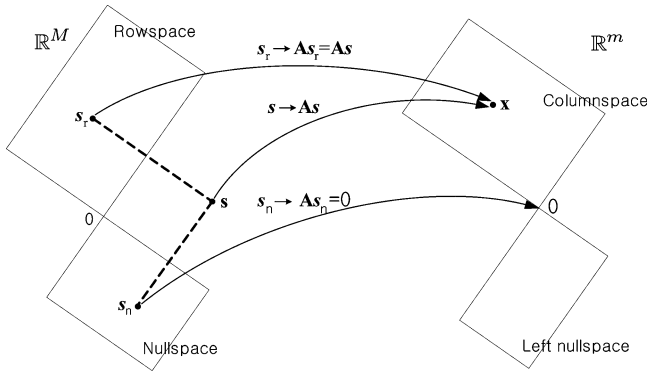


Fig. 2. Representation of $\boldsymbol{s}$ using the four fundamental subspaces of $\mathbf{A}$ and relationships among the subspaces.

IV. Estimation of the Mixing Matrix Based on SSD

The proposed algorithm is based on a three-stage approach as shown in Fig. 1. In a first stage, $\mathbf{A}$ is estimated given $\boldsymbol{x}$. A novel algorithm for estimating $\mathbf{A}$ based on the SSD in the TF domain is presented in this section. Conventional algorithms estimate $\mathbf{A}$ based on clustering algorithms such as the K-means algorithm [19], [29] and on the ratio of the TF transforms [18], [30], [31]. The algorithms in [18] and [29] require that the sources be very sparse in the domain, and the algorithms in [19], [30], and [31] relax this requirement and require that there exist TF regions where only a single source is active for each source. The algorithms in [19], [30] are based on an assumption that there exist many TF points of SSO, and the algorithm in [31] requires there exist at least one small region in the TF plain where there exists only a single source and such a TF region must exist for each source. All aforementioned algorithms require that for each source there exist many TF points of SSO; however, the proposed requires that there exists at least one TF point of SSO and this TF point must exist for each source. The SSD algorithm is less restrictive than the other algorithms.

Before the SSD algorithm is introduced, the STFT of the ith source $s_i[n]$ is defined as

$$S_i[\tau,k] = \sum_{n=-\infty}^{\infty} h[n-\tau]s_i[n]e^{-jkn} \tag{6}$$

at frame τ and frequency bin k where $h[n]$ is a window sequence. Let $\boldsymbol{X}[\tau,k] = \left[X_1[\tau,k] \ \cdots \ X_m[\tau,k]\right]^T$ and $\boldsymbol{S}[\tau,k] = \left[S_1[\tau,k] \ \cdots \ S_M[\tau,k]\right]^T$ denote the STFT coefficients of $\boldsymbol{x}$ and $\boldsymbol{s}$, respectively.

The SSD is based on the ratio of the TF transforms and finds a set of TF points, denoted as $\mathcal{S}_s$, where only a single source is active for each source. It assumes that there exists at least a single TF point of SSO for each source. The main idea of the SSD algorithm is that the ratio of the mixtures in the TF domain is real-valued at the TF point of SSO and complex-valued at the TF point of multiple-source-occupancy (MSO). The mixing matrix is estimated based on the ratio values calculated at the TF points of SSO. The procedure of the SSD is as follows.

1) For a given $\epsilon > 0$, a set $\mathcal{S}_s$ of TF points where only a single source is active is detected for the TF points where both the real and the imaginary parts of the STFT coefficients of the mixtures have sufficient energies such that

$$\mathcal{S}_s = \left\{ [\tau,k] \middle| \left\| \mathrm{Im}\left(\frac{\boldsymbol{X}[\tau,k]}{X_1[\tau,k]} \right) \right\| < \epsilon,\ X_1[\tau,k] \neq 0 \right\} \tag{7}$$

where $\mathrm{Im}(\boldsymbol{x})$ denotes the imaginary part of $\boldsymbol{x}$. We can choose any of the mixtures instead of $X_1[\tau,k]$ as the denominator in (7). For example, if the following is satisfied at $[\tau_s,k_s]$, then $[\tau_s,k_s] \in \mathcal{S}_s$:

$$\begin{aligned} &\left\| \mathrm{Im}\left(\frac{\boldsymbol{X}[\tau_s,k_s]}{X_1[\tau_s,k_s]} \right) \right\| \\ &= \left\| \mathrm{Im}\left(\left[\frac{a_{1i}S_i[\tau_s,k_s]}{a_{1i}S_i[\tau_s,k_s]} \ \frac{a_{2i}S_i[\tau_s,k_s]}{a_{1i}S_i[\tau_s,k_s]} \right.\right.\right. \\ &\quad \left.\left.\left. \cdots \ \frac{a_{mi}S_i[\tau_s,k_s]}{a_{1i}S_i[\tau_s,k_s]} \right]^T \right) \right\| \\ &= \left\| \mathrm{Im}\left(\left[1 \ \frac{a_{2i}}{a_{1i}} \ \cdots \ \frac{a_{mi}}{a_{1i}} \right]^T \right) \right\| = 0 \end{aligned}$$

where a_{li} is the (l,i)th entry of $\mathbf{A}$. This example shows the case that there exists only the ith source at $[\tau_s,k_s]$. On the contrary if both the ith and jth sources, $S_i[\tau,k]$ and $S_j[\tau,k]$, coexist at $[\tau_n,k_n]$, we do not include $[\tau_n,k_n]$ in $\mathcal{S}_s$ because of the following:

$$\begin{aligned} &\left\| \mathrm{Im}\left(\frac{\boldsymbol{X}[\tau_n,k_n]}{X_1[\tau_n,k_n]} \right) \right\| \\ &= \left\| \mathrm{Im}\left(\left[1 \ \frac{a_{2i}S_i[\tau_n,k_n] + a_{2j}S_j[\tau_n,k_n]}{a_{1i}S_i[\tau_n,k_n] + a_{1j}S_j[\tau_n,k_n]} \right.\right.\right. \\ &\quad \left.\left.\left. \cdots \ \frac{a_{mi}S_i[\tau_n,k_n] + a_{mj}S_j[\tau_n,k_n]}{a_{1i}S_i[\tau_n,k_n] + a_{1j}S_j[\tau_n,k_n]} \right]^T \right) \right\| > \epsilon \end{aligned}$$

since $S_i[\tau_n,k_n]$ and $S_j[\tau_n,k_n]$ are complex-valued and can not be erased from numerator and denominator like the previous example.

It should be noted that the value of ϵ needs to be appropriately set. When ϵ is too small, TF points of SSO are difficult to detect, and when ϵ is too large, many unwanted TF points of MSO are detected. In this work, ϵ was experimentally determined.

Fig. 3 illustrates the procedure of the SSD algorithm for the mixtures X_1 and X_2. The TF points colored by white, light gray, and dark gray represent the TF points occupied by no source, a single source, and multiple sources, respectively. As shown in Fig. 3(b), the TF points where both the real and imaginary parts of the STFT coefficients of the mixtures have significant energy are selected. In Fig. 3(c), the ratio of the mixtures selected in (b) is calculated. The ratio values are real at the TF points of SSO and complex at the TF points of MSO. As illustrated in Fig. 3(d), the TF points of SSO are detected by choosing the TF points where the imaginary part of the ratio value is very small. In this example, ϵ_2 is a small positive value.

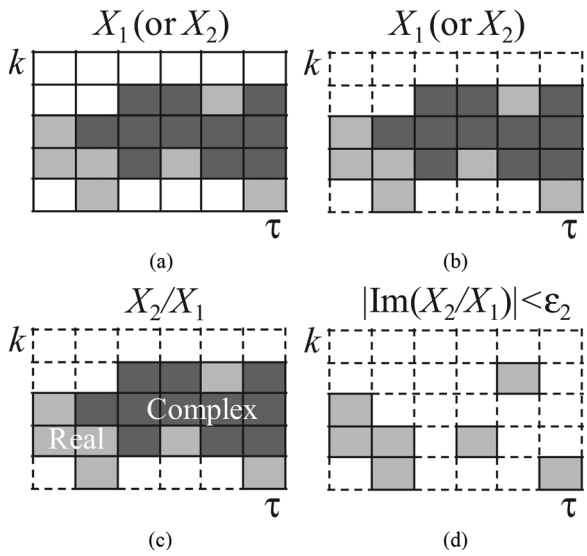


Fig. 3. SSD algorithm is illustrated for the mixtures $X_1[\tau,k]$ and $X_2[\tau,k]$. (a) TF points (tiles) colored by white, light gray, and dark gray represent the TF points occupied by no source, a single source, and multiple sources, respectively. (b) TF points where both the real and imaginary parts of the STFT coefficients of the mixtures have significant energies are selected. (c) Ratio of the mixtures selected in (b) is calculated. The ratio values are real at the TF points of SSO and complex at the TF points of MSO. (d) TF points of SSO are detected by choosing the TF points where the imaginary part of the ratio value is small. In this example, ϵ_2 is a small positive value.

2) The TF points in $\mathcal{S}_s$ are clustered into M classes based on the following ratio vector

$$\mathrm{Re}\left(\frac{\boldsymbol{X}[\tau,k]}{X_1[\tau,k]} \right), \quad \forall [\tau,k] \in \mathcal{S}_s \tag{8}$$

where $\mathrm{Re}(\boldsymbol{x})$ denotes the real part of $\boldsymbol{x}$. We used the K-means algorithm to cluster the TF points in $\mathcal{S}_s$. Any TF point in $\mathcal{S}_s$ where the ith source is active will have the following ratio vector

$$\begin{aligned} \mathrm{Re}\left(\frac{\boldsymbol{X}[\tau,k]}{X_1[\tau,k]} \right) &= \mathrm{Re}\left(\left[1 \ \frac{a_{2i}}{a_{1i}} \ \cdots \ \frac{a_{mi}}{a_{1i}} \right]^T \right) \\ &= \left[1 \ \frac{a_{2i}}{a_{1i}} \ \cdots \ \frac{a_{mi}}{a_{1i}} \right]^T \end{aligned} \tag{9}$$

and the set of the TF points with the above ratio is denoted as $\mathcal{S}_{c_i}$ for $i = 1,\ldots,M$. Using the above algorithm, TF points where only a single source is active are detected for each source.

Finally, we estimate $\mathbf{A}$ using the mixtures in the detected TF region. Prior to the estimation, the real and imaginary parts of all the mixtures are normalized to have unit l_2-norm, respectively.

Using $\boldsymbol{X}[\tau,k]$ in $\mathcal{S}_{c_i}$, the ith column vector of $\mathbf{A}$, denoted as $\boldsymbol{a}_i$, is estimated as follows:

$$\hat{\boldsymbol{a}}_i = \frac{1}{2|\mathcal{S}_{c_i}|} \sum_{[\tau,k]\in\mathcal{S}_{c_i}} \mathrm{Re}(\boldsymbol{X}[\tau,k]) + \mathrm{Im}(\boldsymbol{X}[\tau,k]) \tag{10}$$

where $|\mathcal{S}_{c_i}|$ represents the number of the points in the class for $i = 1,\ldots,M$.

Even when there exists only one TF point for each source, the SSD can accurately detect all points in $\mathcal{S}_s$ and thus estimate the column vectors of $\mathbf{A}$, while many conventional algorithms require many single source TF points to detect the TF points [30], [31].

V. Parameter Estimation Based on EM

In signal processing, the performance of an algorithm is often determined by how well the distribution fits the data. The distributions of multimedia data such as speech, audio, and image differ from one another in both time/spatial and transform domain, and the success of an algorithm is directly linked to how well the algorithm can model the underlying distributions. The distributions have often been modeled using a restrictive class of the distributions such as the Laplacian and the Gaussian. Recently, the student t-distribution [20], [21] and the generalized hyperbolic distribution [22] have been used to model the underlying source distribution. Both distributions are uni-modal distributions. The student t-distribution has one parameter to estimate but can model only the distribution of a sparse source. The generalized hyperbolic distribution can cover a wider range of distributions even with skewness but has five parameters to estimate. A large number of the parameters can increase the degree of freedom of the distribution but can lead to large estimation error and high computational complexity.

In this paper, in order to unmix the mixtures of sources of various distributions, the GGD is used to model both super- and sub-Gaussian distributions by adjusting a single parameter [32]–[34]. It can generate various super- and sub-Gaussian distributions, however, it can model only uni-modal and symmetric distributions. It is given in the following mathematical expression:

$$p_s(s;\sigma,\beta) = \frac{\nu(\beta)}{\sigma} \exp\left\{-c(\beta)\left|\frac{s-\mu}{\sigma}\right|^{2/(1+\beta)}\right\} \tag{11}$$

where

$$c(\beta) = \left(\frac{\Gamma(\frac{3}{2}(1+\beta))}{\Gamma(\frac{1}{2}(1+\beta))}\right)^{1/(1+\beta)} \tag{12}$$

$$\nu(\beta) = \frac{\Gamma(\frac{3}{2}(1+\beta))^{1/2}}{(1+\beta)\Gamma(\frac{1}{2}(1+\beta))^{3/2}} \tag{13}$$

and $\Gamma(\cdot)$ is the Gamma function, which is given as $\Gamma(x) = \int_0^\infty t^{x-1}\exp(-t)\mathrm{d}t$. The parameters $\sigma > 0$ and μ are the standard deviation and the mean of the population, respectively. The parameter $\beta > -1$ can be regarded as a measure of kurtosis indicating the extent of the *non-Gaussianity* of the distribution. The parameter β controls the distance between the GGD and a normal distribution. Fig. 4 shows the shape of the GGD with different values of β when $\sigma = 1$. As shown, when $\beta = 0$, the distribution is a normal distribution. When $\beta = 1$, it is the Laplace distribution. When $\beta = -1$, it is the uniform distribution. In this work, the mean of the source is zero.

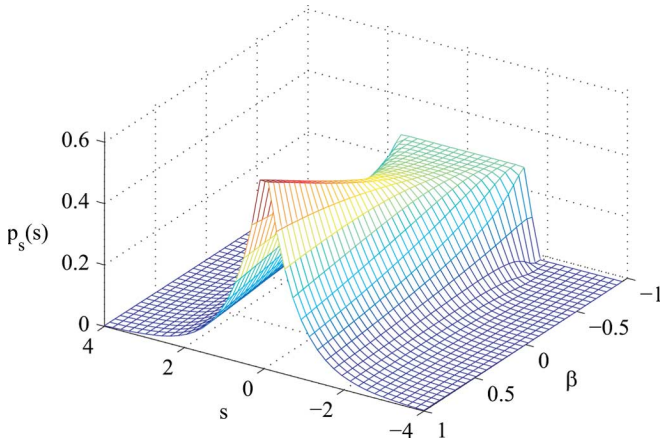


Fig. 4. Shapes of the GGD according to the values of β when each variance equals to one.

In the proposed algorithm, the parameters $\boldsymbol{\theta} = \{\boldsymbol{\beta},\boldsymbol{\sigma}\}$ of the GGD where $\boldsymbol{\beta} = [\beta_1 \ \cdots \ \beta_M]^T$ and $\boldsymbol{\sigma} = [\sigma_1 \ \cdots \ \sigma_M]^T$ are estimated to maximize the likelihood of the incomplete data (mixtures) $\boldsymbol{x}$ using the EM algorithm in [35]. The subscript indicates the index of the source.

Using the latent nullspace component $\boldsymbol{z}$, the log-likelihood of $\boldsymbol{x}$ can be expressed as

$$\log p(\boldsymbol{x}|\boldsymbol{\theta};\mathbf{A}) = \log p(\boldsymbol{z},\boldsymbol{x}|\boldsymbol{\theta};\mathbf{A}) - \log p(\boldsymbol{z}|\boldsymbol{x},\boldsymbol{\theta};\mathbf{A}). \tag{14}$$

Henceforth, $\mathbf{A}$ will be assumed known and omitted from the expression for simplicity. When the conditional expectation of $\log p(\boldsymbol{x}|\boldsymbol{\theta})$ is taken with respect to the posterior probability of $\boldsymbol{z}$, we obtain the auxiliary function $Q(\boldsymbol{\theta},\overline{\boldsymbol{\theta}})$ as follows:

$$Q(\boldsymbol{\theta},\overline{\boldsymbol{\theta}}) = E_{p(\boldsymbol{z}|\boldsymbol{x},\overline{\boldsymbol{\theta}})}[\log p(\boldsymbol{z},\boldsymbol{x}|\boldsymbol{\theta})] \tag{15}$$

where $\overline{\boldsymbol{\theta}}$ represents the updated parameter in the previous iteration, and $E_{p(\boldsymbol{z}|\boldsymbol{x},\overline{\boldsymbol{\theta}})}[f]$ denotes the expectation of function f with respect to $p(\boldsymbol{z}|\boldsymbol{x},\overline{\boldsymbol{\theta}})$. The convergence of the EM algorithm lies in fact that if we choose $\boldsymbol{\theta}$ so that $Q(\boldsymbol{\theta},\overline{\boldsymbol{\theta}}) \geq Q(\overline{\boldsymbol{\theta}},\overline{\boldsymbol{\theta}})$, then $\log p(\boldsymbol{x}|\boldsymbol{\theta}) \geq \log p(\boldsymbol{x}|\overline{\boldsymbol{\theta}})$. Therefore, the parameter $\boldsymbol{\theta}$ is updated at every iteration to maximize $Q(\boldsymbol{\theta},\overline{\boldsymbol{\theta}})$ as follows:

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \ Q(\boldsymbol{\theta},\overline{\boldsymbol{\theta}}) \tag{16}$$

$$= \arg\max_{\boldsymbol{\theta}} \ E_{p(\boldsymbol{z}|\boldsymbol{x},\overline{\boldsymbol{\theta}})}[\log p(\boldsymbol{z},\boldsymbol{x}|\boldsymbol{\theta})]. \tag{17}$$

Since the calculation to obtain the expectation of (17) is generally intractable, it is approximated using the Monte Carlo integration [36], [37]. That is, (17) can be approximated as

$$\hat{\boldsymbol{\theta}} \approx \arg\max_{\boldsymbol{\theta}} \frac{1}{K}\sum_{l=1}^{K} \log p(\boldsymbol{z}^{(l)},\boldsymbol{x}|\boldsymbol{\theta}) \tag{18}$$

where $\{\boldsymbol{z}^{(1)} \cdots \boldsymbol{z}^{(K)}\}$ are the K samples drawn from $p(\boldsymbol{z}|\boldsymbol{x}, \overline{\boldsymbol{\theta}})$ using the Markov chain Monte Carlo method. Since the representation in (4) can be rewritten as

$$\boldsymbol{s} = [\mathbf{A}^{\dagger},\ \mathbf{V}] \begin{bmatrix} \boldsymbol{x} \\ \boldsymbol{z} \end{bmatrix} \tag{19}$$

where $[\mathbf{A}^{\dagger},\ \mathbf{V}]$ is an $M \times M$ matrix, (18) can be expressed as

$$\hat{\boldsymbol{\theta}} \approx \arg\max_{\boldsymbol{\theta}} \frac{1}{K} \sum_{l=1}^{K} \log \left| \det[\mathbf{A}^{\dagger}, \mathbf{V}] \right| p_{\boldsymbol{s}}(\boldsymbol{s}^{(l)}|\boldsymbol{\theta}) \tag{20}$$

$$\approx \arg\max_{\boldsymbol{\theta}} \frac{1}{K} \sum_{l=1}^{K} \log p_{\boldsymbol{s}}(\boldsymbol{s}^{(l)}|\boldsymbol{\theta}) \tag{21}$$

where $p_{\boldsymbol{s}}(\boldsymbol{s}^{(l)}|\boldsymbol{\theta})$ is the GGD parametric form, and $\left|\det[\mathbf{A}^{\dagger}, \mathbf{V}]\right|$ is the absolute value of the determinant of $[\mathbf{A}^{\dagger}, \mathbf{V}]$. Finally, the model parameters are updated at every iteration by maximizing the log-likelihood of the sources obtained from $\{\boldsymbol{z}^{(1)} \cdots \boldsymbol{z}^{(K)}\}$ that is sampled from $p(\boldsymbol{z}|\boldsymbol{x}, \overline{\boldsymbol{\theta}})$.

VI. LNCE Based on MMSE

This section shows how the sources are estimated based on the MMSE criterion. Once $\boldsymbol{z}$ is estimated, $\hat{\boldsymbol{s}}$ can be obtained directly from (4). Let $\tilde{\boldsymbol{x}}$ denote the observation that consists of $\boldsymbol{x}$ and $\hat{\boldsymbol{\theta}}$, that is, $\tilde{\boldsymbol{x}} = \{\boldsymbol{x}, \hat{\boldsymbol{\theta}}\}$. Let $C(z, \hat{z}(\tilde{\boldsymbol{x}}))$ denote a cost function that assigns a cost to a pair of actual latent component value and its estimate where $\hat{z}(\tilde{\boldsymbol{x}})$ represents the estimate of $\boldsymbol{z}$ when $\tilde{\boldsymbol{x}}$ is given. The Bayes estimator of $\boldsymbol{z}$ is given as

$$\hat{\boldsymbol{z}}_B = \arg\min_{\hat{\boldsymbol{z}}} \int_{-\infty}^{\infty} C(z, \hat{\boldsymbol{z}}(\tilde{\boldsymbol{x}})) p(\boldsymbol{z}|\tilde{\boldsymbol{x}}) d\boldsymbol{z} \tag{22}$$

where $\hat{\boldsymbol{z}}_B$ is an optimum estimator for C when $\tilde{\boldsymbol{x}}$ is given [38].

One can generate an infinite number of different estimators depending on the choice of $C(\boldsymbol{z}, \hat{\boldsymbol{z}}(\tilde{\boldsymbol{x}}))$. In the proposed algorithm, we aim to estimate the sources such that the MSE of the estimation is minimized in order to maximize the signal-to-interference ratio (SIR), which is defined in (28). When $\mathbf{A}$ is known, the square error of $\hat{\boldsymbol{s}}$ can be represented as

$$\begin{aligned} ||\boldsymbol{s} - \hat{\boldsymbol{s}}||^2 &= ||\boldsymbol{s}_r + \boldsymbol{s}_n - (\boldsymbol{s}_r + \hat{\boldsymbol{s}}_n)||^2 \\ &= ||\boldsymbol{s}_n - \hat{\boldsymbol{s}}_n||^2 \\ &= ||\mathbf{V}\boldsymbol{z} - \mathbf{V}\hat{\boldsymbol{z}}||^2 \\ &= (\boldsymbol{z} - \hat{\boldsymbol{z}})^T \mathbf{V}^T \mathbf{V} (\boldsymbol{z} - \hat{\boldsymbol{z}}) \\ &= ||\boldsymbol{z} - \hat{\boldsymbol{z}}||^2 \end{aligned} \tag{23}$$

where $\hat{\boldsymbol{s}}_n$ and $||\boldsymbol{s}||$ represent an estimate of $\boldsymbol{s}_n$ and the l_2-norm of $\boldsymbol{s}$, respectively, and $\mathbf{V}^T\mathbf{V} = \mathbf{I}$. From (23), we notice that the MSE of the source estimation is equal to that of the LNCE. Therefore, the cost function $C(\boldsymbol{z}, \hat{\boldsymbol{z}}(\tilde{\boldsymbol{x}}))$ is chosen to be the MSE cost function as follows:

$$C_{\mathrm{MS}}(z, \hat{\boldsymbol{z}}(\tilde{\boldsymbol{x}})) = ||\boldsymbol{z} - \hat{\boldsymbol{z}}||^2. \tag{24}$$

The Bayes estimator for this cost function is the posterior mean given by

$$\hat{\boldsymbol{z}}_{\mathrm{MS}} = \int_{-\infty}^{\infty} \boldsymbol{z} p(\boldsymbol{z}|\tilde{\boldsymbol{x}}) d\boldsymbol{z} \tag{25}$$

which can be obtained by sampling from $p(\boldsymbol{z}|\tilde{\boldsymbol{x}})$ as

$$\hat{\boldsymbol{z}}_{\mathrm{MS}} \approx \frac{1}{K} \sum_{l=1}^{K} \boldsymbol{z}^{(l)} \tag{26}$$

where $\{\boldsymbol{z}^{(1)} \cdots \boldsymbol{z}^{(K)}\}$ are the K drawn samples from $p(\boldsymbol{z}|\tilde{\boldsymbol{x}})$, that is, $p(\boldsymbol{z}|\boldsymbol{x}, \hat{\boldsymbol{\theta}})$. The drawn samples come from the same probability function $p(\boldsymbol{z}|\boldsymbol{x}, \hat{\boldsymbol{\theta}})$ to approximate the auxiliary function in (18). Using $\hat{\boldsymbol{z}}_{\mathrm{MS}}$, the sources satisfying the MSE criterion, $\hat{\boldsymbol{s}}_{\mathrm{MS}}$, can be expressed as

$$\hat{\boldsymbol{s}}_{\mathrm{MS}} = \mathbf{A}^{\dagger}\boldsymbol{x} + \mathbf{V}\hat{\boldsymbol{z}}_{\mathrm{MS}}. \tag{27}$$

VII. Simulations

Simulations were performed on both synthetically generated signals and speech/audio signals using the proposed and conventional algorithms. In the simulations, sparse (super-Gaussian) and nonsparse (sub-Gaussian) signals were separated from the underdetermined mixtures. First, in case of synthetic signals, the proposed algorithm was applied in the time domain with an exact mixing matrix $\mathbf{A}$. Separations were performed on two sets of the sources: sources that consisted of only sparse sources and that consisted of both sparse and nonsparse signals. Next, speech and audio signals were separated in the TF domain. Prior to the separation, $\mathbf{A}$ was estimated using the SSD algorithm, and the estimated $\hat{\mathbf{A}}$ was used in the separation simulation. A simulation to separate signals in a noise environment was also performed.

A. Separation of Synthetic Signals

1) Sparse Signals: The separation of M synthetic sparse signals from $m = 3$ mixtures was performed in the time domain for $M = 4, 5$, and 6. In this simulation, the mixing matrix $\mathbf{A}$ was known. The proposed algorithm was compared to the l_1-norm minimization [14] and the FOCUSS algorithms [39] when $p = 0.8$. The simulation settings were as follows. Synthetic sparse signals were generated by generating 1000 Gaussian samples using *randn* command of Matlab and substituting 90% of the samples chosen randomly by zeros for each source. The mixing matrix $\mathbf{A}$ whose rank is 3 was also generated randomly every simulation, and all the columns of $\mathbf{A}$ were normalized to have unit l_2-norm. The parameter $\boldsymbol{\beta}$ and $\boldsymbol{\sigma}^2$ of the GGD were initialized randomly using the uniform function of the range $[-1,\ 9]$ and to value 1 for all sources, respectively. Under these conditions, 50 simulations were performed for each M, and the performance was evaluated in terms of the averaged SIR over 50

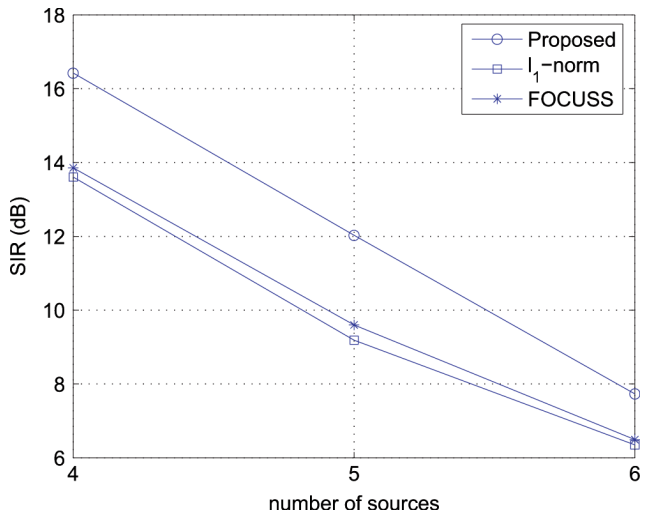


Fig. 5. Performance of estimating $\boldsymbol{s}$ according to the number of underlying sources from three mixtures.

simulations. Given an original source s_i and its estimate $\hat{s}_i$, SIR in decibels is defined as

$$\mathrm{SIR} = 10\log_{10}\frac{\sum_{i=1}^{M} E\{s_i^2\}}{\sum_{i=1}^{M} E\{(s_i-\hat{s}_i)^2\}}. \tag{28}$$

Fig. 5 illustrates the averaged SIRs when the number of the sources increase from $M = 4$ to 6. The proposed algorithm achieved about 2.5 dB higher SIR for $M = 4$ and $M = 5$ and 1.2 dB higher SIR for $M = 6$ than the other algorithms, respectively. The SIRs obtained using the FOCUSS algorithm were slightly higher than those obtained using the l_1-norm minimization. The performance was degraded as the number of the underlying sources increased.

Fig. 6 shows the learning curves of the proposed algorithm for updating the parameters of GGD. The solid and dotted lines represent the parameters estimated using the proposed algorithm and from the original sources, respectively. As shown in Fig. 6(a), $\boldsymbol{\beta}$ converged to a value close to those obtained from the original sources that correspond to super-Gaussian distributions from randomly initialized values after iteration 2, and as shown in Fig. 6(b), $\boldsymbol{\sigma}^2$ also converged to the values close to those estimated from the original sources after iterations 2.

2) Sparse and Non-Sparse Signals: The separation of four synthetically generated sparse (super-Gaussian) and nonsparse (sub-Gaussian) sources from three mixtures was performed in the time domain using the proposed and l_1-norm minimization algorithms. In this simulation, $\mathbf{A}$ was known.

Fig. 7 shows separation results obtained using both the proposed and the l_1-norm minimization algorithms. The first two plots in Fig. 7(a) are the sub-Gaussian sources, and the other two are the super-Gaussian sources. The three plots in Fig. 7(b) represent the three mixtures. The four plots in Figs. 7(c) and 7(d) represent the estimated sources using the proposed and the l_1-norm minimization algorithms, respectively.

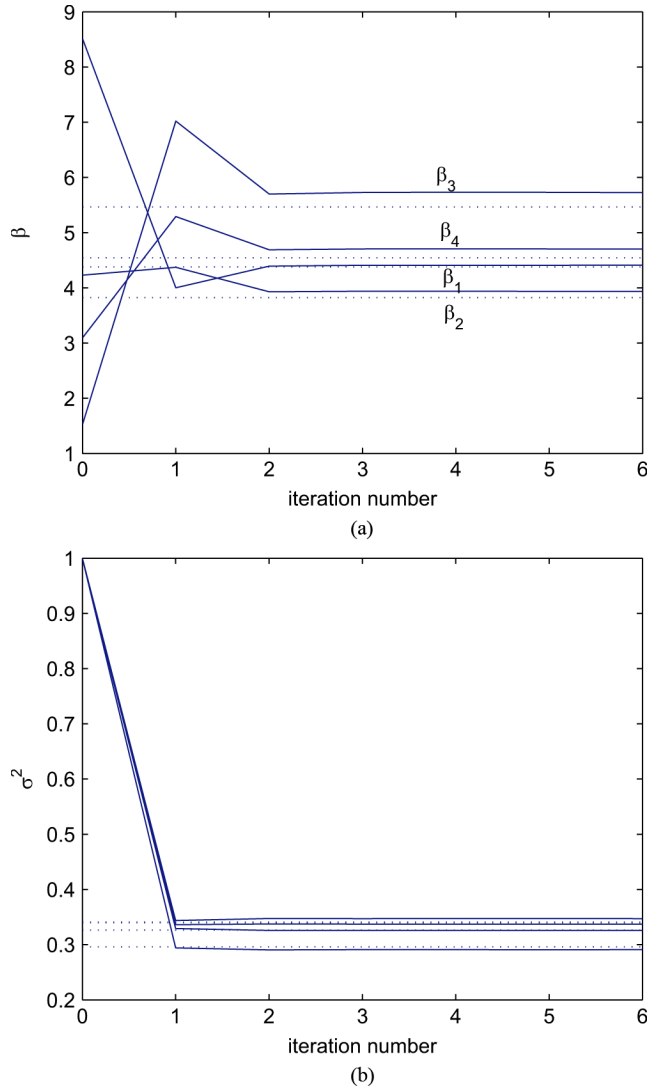


Fig. 6. Learning curves of the proposed algorithm for updating the parameters of the GGD based on the EM. The solid and dotted lines represent the parameters estimated using the proposed algorithm and from the original sources, respectively. (a) Learning curves for $\boldsymbol{\beta}$ of GGD. (b) Learning curves for $\boldsymbol{\sigma}^2$ of GGD.

As shown, the proposed algorithm separated both super- and sub-Gaussian sources better than the l_1-norm minimization algorithm. This result can also be verified in Table I, which shows the kurtosis values of the original and estimated sources: κ_i^o, κ_i^p, and κ_i^1 are the kurtosis values of the ith original source, estimated source using the proposed algorithm, and estimated source using the l_1-norm minimization algorithm for $i = 1$, 2, 3, and 4, respectively. As shown in the table, the proposed algorithm estimated the sub-Gaussian sources as well as the super-Gaussian sources. The proposed and l_1-norm minimization algorithms achieved 8.78 and 3.56 dB SIR, respectively.

B. Separation of Speech and Audio Signals

Simulation to separate speech and audio sources from underdetermined mixtures was performed. For speech and audio signals, a sparser representation can be obtained in the TF domain. In order to achieve better separation of the signals, the proposed underdetermined BSS algorithm was applied in the TF domain

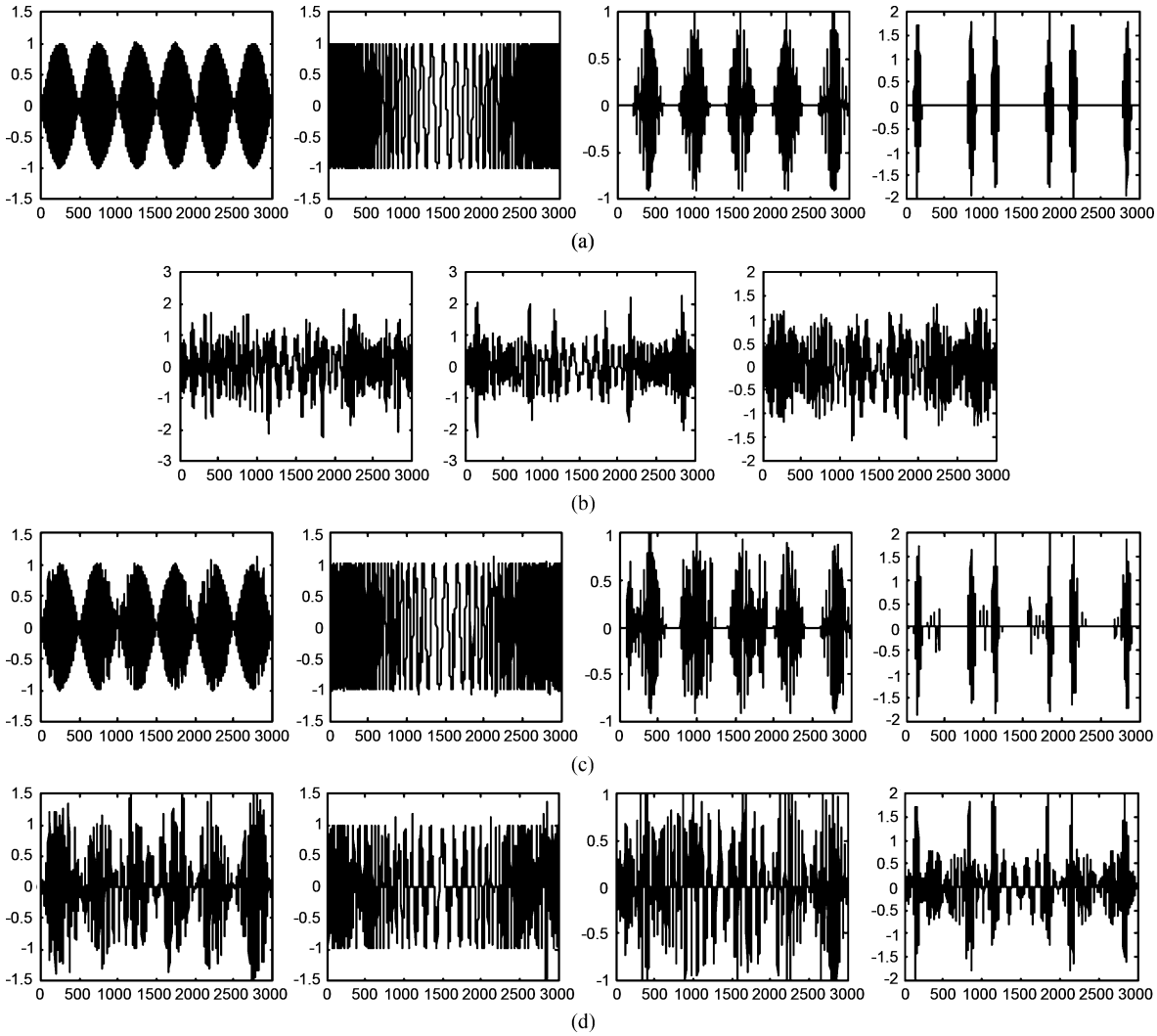


Fig. 7. Separation of the synthetically generated two sub- and two super-Gaussian sources from three mixtures. (a) Four original sub- and super-Gaussian sources. (b) Three mixtures. (c) Four estimated sources using the proposed algorithm. (d) Four estimated sources using the l_1-norm minimization algorithm.

TABLE I
KURTOSIS VALUES OF THE ORIGINAL SOURCES AND THE ESTIMATED SOURCES USING THE PROPOSED AND THE l_1-NORM MINIMIZATION ALGORITHMS

i	1	2	3	4
κ_i^o	-0.7504	-1.5006	1.0561	8.5304
κ_i^p	-0.4707	-1.3752	3.2661	8.5870
κ_i^1	0.5825	-0.4632	1.2694	5.7405

with the transform distribution of the sources modeled by the GGD. For the TF transform, we used the STFT with the Hanning window whose length was 512 samples and 50% overlap.

In this simulation, the separation was performed on the four sets of the signals: various instruments such as guitar, drum, base, and keyboard (set 1), speech signals (set 2), various kinds of sounds such as bell, car, airplane, and train (set 3), and various genres of music such as classic, rock, jazz, and pop (set 4). All the signals were sampled at 16 kHz and had the length of 80 000 samples except that speech signals had the length of 60 000 samples. The proposed algorithm was compared to the conventional algorithms based on the TF nondisjoint assumption [19], the TF disjoint assumption [18], and the l_1-norm minimization [14]. The l_1-norm minimization algorithm was applied in the TF domain, and the transform distribution of the sources was modeled by a Laplacian. A modified version of the algorithm proposed in [18] that can deal with more than two mixtures was implemented for comparison [19]. Prior to the separation simulation, $\mathbf{A}$ was estimated using aforementioned algorithms in the TF domain. The l_1-norm minimization algorithm used an estimate using the SSD algorithm.

1) Estimation of $\mathbf{A}$*:* A mixing matrix $\mathbf{A}$ was estimated using the proposed and conventional algorithms in [18], [19]. In the evaluation, the columns of $\mathbf{A}$ were all normalized. Typically, ϵ was set to 0.01.

For example, the result of estimating 3×4 mixing matrix $\mathbf{A}$ using the proposed algorithm with the instrument signals is shown. We obtained 3 mixtures by mixing 4 sources using 3×4 mixing matrix $\mathbf{A}$ given as

$$\mathbf{A} = \begin{bmatrix} 0.6547 & 0.6516 & 0.8830 & -0.5571 \\ 0.3780 & -0.5923 & 0.3532 & 0.7428 \\ -0.6547 & 0.4739 & 0.3091 & 0.3714 \end{bmatrix}. \quad (29)$$

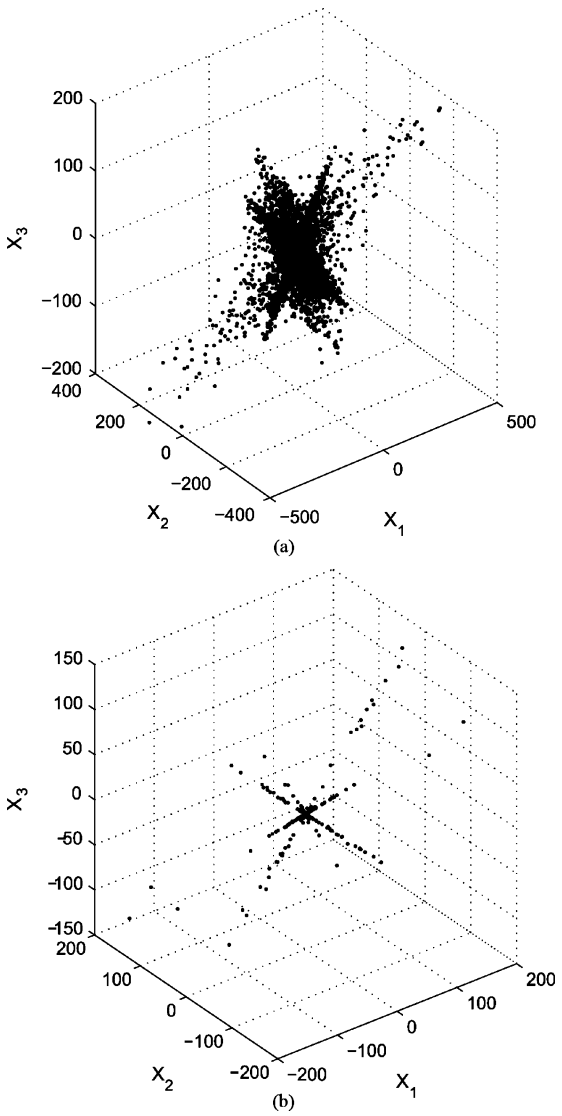


Fig. 8. Three-dimensional scatter plots of the mixtures in the TF domain. (a) Scatter plot of the real and imaginary parts of the STFT coefficients of the mixtures in $\mathcal{S}$. (b) Scatter plot of the real and imaginary parts of the STFT coefficients of the mixtures in $\mathcal{S}_s$.

The scatter plots of the real and imaginary parts of the STFT coefficients of the mixtures in $\mathcal{S}$ and $\mathcal{S}_s$ are shown in Fig. 8. The 3-D scatter plot of the STFT coefficients of the mixtures in $\mathcal{S}$ is shown in Fig. 8(a). It shows a single big cloud, so it is difficult to find the column vectors from the scatter plot. Fig. 8(b) presents the 3-D scatter plot of the STFT coefficients of the mixtures in $\mathcal{S}_s$. The data points are plotted along the directions of the column vectors of $\mathbf{A}$, which leads to a good estimation of $\mathbf{A}$. The estimated matrix $\hat{\mathbf{A}}$ using the proposed algorithm based on the SSD is as follows:

$$\hat{\mathbf{A}} = \begin{bmatrix} 0.6504 & 0.6528 & 0.8820 & -0.5735 \\ 0.3818 & -0.5921 & 0.3664 & 0.7279 \\ -0.6567 & 0.4726 & 0.2965 & 0.3757 \end{bmatrix} \tag{30}$$

after reordering and changing the sign. All the column vectors of $\hat{\mathbf{A}}$ were estimated to be close to those of $\mathbf{A}$.

The $3 \times M$ mixing matrix $\mathbf{A}$ was estimated with the speech signals using the proposed and conventional algorithms in [18], [19] for $M = 4$, 5, and 6. The performance of the estimation is evaluated in terms of a new criterion, denoted as $\varepsilon_{\mathbf{A}}$, which is defined as

$$\varepsilon_{\mathbf{A}} = \frac{1}{M}\|\mathbf{A} - \hat{\mathbf{A}}\|_F^2 \tag{31}$$

where $\|\mathbf{A}\|_F$ denotes the Frobenius norm of $\mathbf{A}$. Fig. 9 shows the averaged $\epsilon_{\mathbf{A}}$ for 50 simulations according to M. A $3 \times M$ mixing matrix $\mathbf{A}$ whose condition number is less than 5 was randomly generated every simulation for $M = 4$, 5, and 6, and the columns of $\mathbf{A}$ were all normalized.

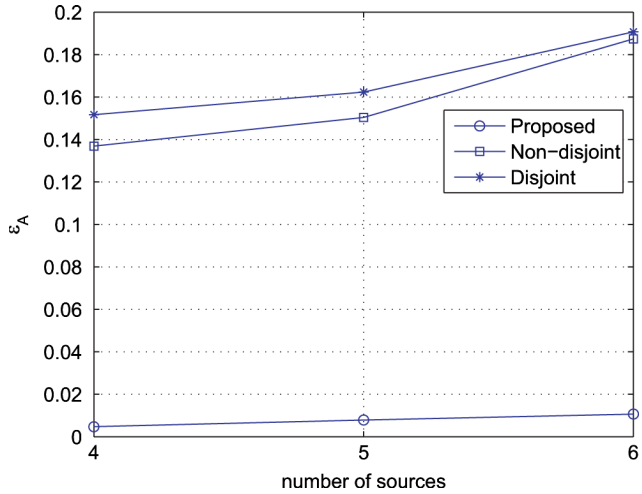


Fig. 9. Performance of estimating $\mathbf{A}$ according to the number of the underlying sources using three mixtures.

As shown in the figure, the proposed algorithm based on SSD performed better than both algorithms. As the number of the underlying sources increased, the estimation error also increased; the increment of the estimation error of the proposed algorithm was much smaller than those of the other algorithms.

Simulation with additive white Gaussian noise was performed by varying signal-to-noise ratio (SNR) from 5 to 35 dB in the 3×4 case. In this simulation, SNR is defined as the ratio of the power of each mixture to that of the noise. For a reliable estimation of $\mathbf{A}$, the proposed used only high energy TF points in determining $\mathbf{A}$. This is possible since using the SSD algorithm we only need at least one TF point of SSO for each source. Fig. 10 shows the performances of various mixing matrix estimation algorithms under various SNR conditions. As shown in the figure, the proposed algorithm estimated $\mathbf{A}$ more accurately and was more robust to noise than the other algorithms. When the SNR was above 10 dB, the proposed algorithm was relatively insensitive to noise, while the performances of the other algorithms degraded linearly with the SNR; however, when the SNR is below 10 dB, the performance of the proposed algorithm degraded faster than the other algorithms. This can be attributed to the difficulty in detecting TF points of SSO at low SNR.

2) Estimation of $\mathbf{s}$*:* Simulations to separate four sources from three mixtures were performed on the four sets of the sources in the TF domain. Fig. 11 shows the SIR performances for each set. The estimated matrices $\hat{\mathbf{A}}$ that were obtained using each algorithm were used in the subsequent simulation for separating

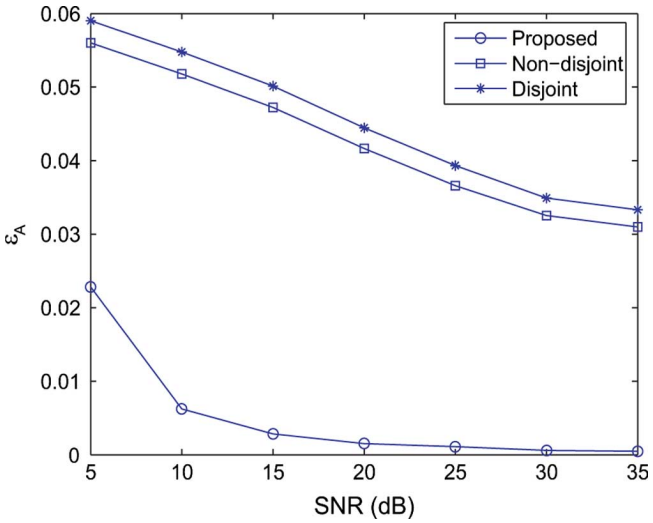


Fig. 10. Estimation error of $\mathbf{A}$ according to SNR.

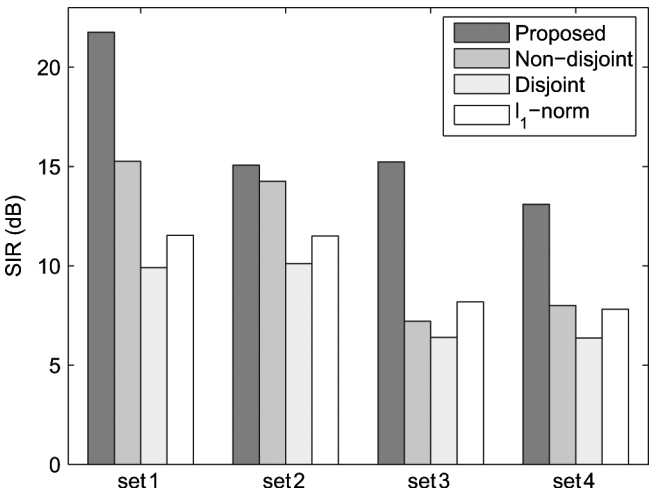


Fig. 11. Performances of the proposed, the TF nondisjoint, the TF disjoint, and the l_1-norm minimization algorithms to separate four sources from three mixtures for each set: instruments (set 1), speech (set 2), various sounds (set 3), and music (set 4).

the sources. The l_1-norm minimization algorithm used $\hat{\mathbf{A}}$ estimated using the SSD algorithm. As shown in the figure, the proposed algorithm achieved more than 6.5 dB improvement over others in terms of SIR for set 1, 3, and 4 and more than 1-dB improvement over others in terms of SIR for set 2. The algorithms based on the TF disjoint and nondisjoint led to sources with audible distortion. The demo files are available at http://mmp.kaist.ac.kr/~ifree/demo.html.

By varying the SNR from 5 to 35 dB, simulation to separate four speech signals from three mixtures was performed. Fig. 12 shows the SIR performance of each algorithm according to the SNR. The proposed achieved higher SIR than the other algorithms when the SNR was above 10 dB. When the SNR was below 10 dB, the algorithm based on the TF disjoint achieved higher SNR than the other algorithms, however, the audible distortions in the estimates of the sources were much more severe than those of the other algorithms.

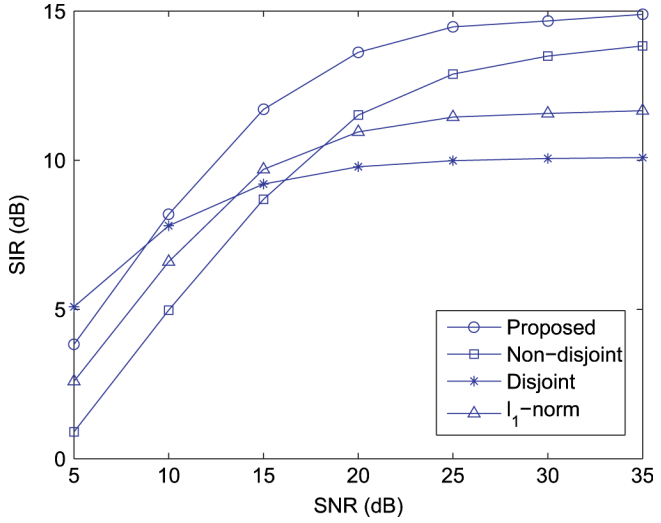


Fig. 12. Performances of the proposed, the TF nondisjoint, the TF disjoint, and the l_1-norm minimization algorithms to separate four sources from three mixtures according to SNR for speech.

VIII. Conclusion

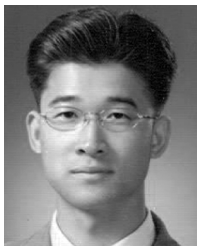
This paper considers a problem of blindly separating the sources of sub- and super-Gaussian distributions from underdetermined mixtures. In this paper, first a novel algorithm for estimating $\mathbf{A}$ based on the SSD in the TF domain is proposed. The TF points that include a single source are detected using the SSD algorithm, and the TF coefficients of the mixtures at the TF points are used to estimate $\mathbf{A}$. Even when there is only a single TF point where the single source is active, the proposed algorithm can detect the TF point and estimate the column vector of $\mathbf{A}$ that corresponds to the source active at the detected TF point. Second, in order to characterize and model sub- and super-Gaussian distributions, the parametric GGD is used, and its parameters are estimated based on the EM algorithm. Finally, $\boldsymbol{z}$ pertaining to estimating $\boldsymbol{s}$ is estimated to minimize the MSE between its true and estimated value. This reduces the dimension of the estimator from M of $\boldsymbol{s}$ to $M-m$ of $\boldsymbol{z}$. This paper showed that the estimation of $\boldsymbol{z}$ based on the MMSE leads to an optimum estimation of $\boldsymbol{s}$ for the MSE criterion.

Simulation results show that the proposed underdetermined BSS algorithm estimated the mixing matrix with higher accuracy and separated both sparse and nonsparse signals with higher SIR than the conventional algorithms.

References

[1] S. Amari and A. Cichocki, "Adaptive blind signal processing. neural network approaches," *Proc. IEEE*, vol. 86, pp. 2026–2048, Oct. 1998.
[2] A. Mansour, A. K. Barros, and N. Ohnishi, "Blind separation of sources: Methods, assumptions and applications," *IEICE Trans. Fund. Electron., Commun., Comput. Sci.*, vol. E83-A, pp. 1498–1512, 2000.
[3] A. Cichocki and S. Amari, *Adaptive Blind Signal and Image Processing*. New York: Wiley, 2003.
[4] T. W. Lee, *Independent Component Analysis*. Norwell, MA: Kluwer, 1998.
[5] A. J. Bell and T. J. Sejnowski, "An information-maximisation approach to blind separation and blind deconvolution," *Neural Comput.*, vol. 7, no. 6, pp. 1129–1159, 1995.
[6] J. F. Cardoso, "Blind signal processing: Statistical principles," *Proc. IEEE*, pp. 2009–2025, 1998.

[7] A. Belouchrani, K. Abed-Meraim, J. F. Cardoso, and E. Moulines, "A blind source separation technique using second-order statistics," *IEEE Trans. Signal Process.*, vol. 45, no. 2, pp. 434–444, Feb/ 1997.
[8] S. Degerine and R. Malki, "Second-order blind separation of sources based on canonical partial innovations," *IEEE Trans. Signal Process.*, vol. 48, no. 3, pp. 629–641, Mar. 2000.
[9] S. Choi and A. Cichocki, "Blind separation of non-stationary sources in noisy mixtures," *Electron. Lett.*, vol. 36, no. 9, pp. 849–850, 2000.
[10] D. T. Pham and J. F. Cardoso, "Blind separation of instantaneous mixtures of non-stationary sources," *IEEE Trans. Signal Process.*, vol. 49, no. 9, pp. 1837–1848, Sep. 2001.
[11] S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Sci. Comput.*, vol. 20, no. 1, pp. 33–61, 1998.
[12] D. L. Donoho and M. Elad, "Maximal sparsity representation via l_1 minimization," *Proc. Nat. Acad. Sci.*, vol. 100, pp. 2197–2202, 2003.
[13] M. S. Lewicki and T. J. Sejnowski, "Learning overcomplete representations," *Neural Comput.*, vol. 12, pp. 337–365, 2000.
[14] T. W. Lee, M. S. Lewicki, M. Girolami, and T. J. Sejnowski, "Blind source separation of more sources than mixtures using overcomplete representation," *IEEE Signal Process. Lett*, vol. 6, no. 4, pp. 87–90, 1999.
[15] M. Zibulevsky and B. A. Pearlmutter, "Blind source separation by sparse decomposition," *Neural Comput.*, vol. 13, no. 4, pp. 863–882, 2001.
[16] M. Girolami, "A variational method for learning sparse and overcomplete representations," *Neural Comput.*, vol. 13, pp. 2517–2532, 2001.
[17] Y. Li, A. Cichocki, and S. Amari, "Sparse component analysis for blind source separation with less sensors than sources," in *Proc. 4th Int. Symp. Independent Component Analysis Blind Signal Separation*, 2003, pp. 89–94.
[18] O. Yilmaz and S. Rickard, "Blind separation of speech mixtures via time-frequency masking," *IEEE Trans. Signal Process.*, vol. 52, no. 7, pp. 1830–1847, Jul. 2004.
[19] A. Aissa-El-Bey, N. Linh-Trung, K. Abed-Meraim, A. Belouchrani, and Y. Grenier, "Underdetermined blind separation of nondisjoint sources in the time-frequency domain," *IEEE Trans. Signal Process.*, vol. 55, no. 3, pp. 897–907, Mar. 2007.
[20] A. T. Cemgil, C. Fevotte, and S. J. Godsill, "Variational and stochastic inference for Bayesian source separation," *Digit. Signal Process.*, vol. 17, pp. 891–913, 2007.
[21] A. T. Cemgil, C. Fevotte, and S. J. Godsill, "Blind separation of sparse sources using variational EM," in *Proc. 13th Eur. Signal Process. Conf.*, 2005.
[22] H. Snoussi and J. Idier, "Bayesian blind separation of generalized hyperbolic processes in noisy and underdetermined mixtures," *IEEE Trans. Signal Process.*, vol. 54, no. 9, pp. 3257–3269, Sep. 2006.
[23] S. G. Kim and C. D. Yoo, "Underdetermined independent component analysis by data generation," in *Proc. ICA*, Granada, Spain, 2004, pp. 445–452.
[24] S. G. Kim and C. D. Yoo, "Underdetermined blind source separation based on generalized Gaussian distribution," presented at the IEEE Int. Workshop MLSP, Maynooth, Ireland, 2006.
[25] C. R. Rao and S. K. Mitra, *Generalized Inverse of Matrices and its Applications*. New York: Wiley, 1971.
[26] D. Donoho, "The simplest solution to an underdetermined system of linear equations," in *Proc. IEEE Int. Symp. Inform. Theory*, 2006, pp. 1924–1928.
[27] G. Strang, *Introduction to Linear Algebra*. Cambridge, MA: Wellesley-Cambridge, 2003.
[28] T. K. Moon and W. C. Stirling, *Mathematical Methods and Algorithms for Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 2000.
[29] Y. Li, A. Cichocki, and S. Amari, "Analysis of sparse representation and blind source separation," *Neural Comput.*, vol. 16, pp. 1193–1234, 2004.
[30] Y. Li, S. Amari, A. Cichocki, D. W. C. Ho, and S. Xie, "Underdetermined blind source separation based on sparse representation," *IEEE Trans. Signal Process.*, vol. 54, no. 2, pp. 423–437, Feb. 2006.
[31] F. Abrard and Y. Deville, "A time-frequency blind signal separation method applicable to underdetermined mixtures of dependent sources," *Signal Process.*, vol. 85, no. 7, pp. 1389–1403, 2005.
[32] S. Choi, A. Cichocki, and S. Amari, "Flexible independent component analysis," *J. VLSI Signal Process*, vol. 26, pp. 25–38, 2000.
[33] G. E. P. Box and G. C. Tiao, *Bayesian Inference in Statistical Analysis*. Reading, MA: Addison-Wesley, 1973.
[34] K. Kokkinakis and A. k. Nandi, "Exponent parameter estimation for generalized Gaussian probability density functions with application to speech modeling," *Signal Process.*, vol. 85, pp. 1852–1858, 2005.
[35] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via EM algorithm," *J. Roy. Stat. Soc.*, vol. 39, no. Ser. B, pp. 1–38, 1977.
[36] G. C. Wei and M. A. Tanner, "A Monte Carlo implementation of the EM algorithm and the poor man's data augmentation algorithms," *J. Amer. Stat. Assoc.*, vol. 85, no. 411, pp. 699–704, 1990.
[37] G. S. Fishman, *Monte Carlo: Concepts, Algorithms, and Applications*. New York: Springer-Verlag, 1995.
[38] J. L. Melsa and D. L. Cohn, *Decision and Estimation Theory*. New York: McGraw-Hill, 1978.
[39] I. F. Gorodnitsky and B. D. Rao, "Sparse signal reconstruction from limited data using FOCUSS: A re-weighted minimum norm algorithm," *IEEE Trans. Signal Process*, vol. 45, no. 3, pp. 600–616, Mar. 1997.



SangGyun Kim (S'02–M'08) received the B.S., M.S., and Ph.D. degrees in electrical engineering and computer science (EECS) from Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Korea, in 1999, 2001, and 2008, respectively.

He joined the Multimedia Processing Laboratory in EECS at KAIST in 2001. From March 2008 to December 2008, he was with the School of Information Technology, KAIST, as a Postdoctoral Researcher. He is currently a Postdoctoral Researcher at the Department of Brain and Cognitive Sciences (BCS), Massachusetts Institute of Technology (MIT), Cambridge, MA. His research interests are in the application of statistical signal processing and machine learning to neural and multimedia signals.

Dr. Kim is a member of the IEEE, the Institute of Electrical Engineers of Korea (IEEK), and the Acoustical Society of Korea (ASK).

Chang D. Yoo (S'92–M'96) received the B.S. degree in engineering and applied science from the California Institute of Technology, Pasadena, in 1986, the M.S. degree in electrical engineering from Cornell University, Ithaca, NY, in 1988, and the Ph.D. degree in electrical engineering from the Massachusetts Institute of Technology (MIT), Cambridge, in 1996.

From January 1997 to March 1999, he worked at Korea Telecom as a Senior Researcher. He joined the Department of Electrical Engineering at Korea Advanced Institute of Science and Technology in April 1999. From March 2005 to March 2006, he was with Research Laboratory of Electronics at MIT. His current research interests are in the application of machine learning and digital signal processing in multimedia.

Dr. Yoo is a member of Tau Beta Pi and Sigma Xi.