UNDERDETERMINED BLIND SOURCE SEPARATION BASED ON GENERALIZED GAUSSIAN DISTRIBUTION

SangGyun Kim and Chang D. Yoo

Dept. of EECS, Div. of EE, KAIST
373-1 Guseong-dong, Yuseong-gu, Daejeon 305-701, Republic of Korea
zom@eeinfo.kaist.ac.kr and cdyoo@ee.kaist.ac.kr

ABSTRACT

In this paper, a novel method for separating underlying sources with both sub- and super-Gaussian distributions from the underdetermined mixtures is proposed. The generalized Gaussian distribution (GGD) is used to model simultaneously both sub- and super-Gaussian distributions. The process of finding the most probable decomposition of the mixtures based on the GGD leads to that of minimizing the $L_{\mathbf{p}}$-norm of the estimated sources. The switching condition for determining the decay rate of the GGD is determined by the sign of the kurtosis of the inferred source. In our simulation, the proposed algorithm separated both the sub- and super-Gaussian sources from the underdetermined mixtures and achieved about 1 dB improvement in signal-to-interference (SIR) over the l_1-norm minimization algorithm in separating three speech sources from two mixtures.

1. INTRODUCTION

A blind source separation (BSS) algorithm aims to recover unobserved sources from a number of observed mixtures. The problem that it is solving can be formulated statistically as follows: given M-dimensional random variable vector $\mathbf{x} = [x_1, ..., x_M]^T$ that arises from linear combination of the mutually independent components of N-dimensional unknown random variable $\mathbf{s} = [s_1, ..., s_N]^T$, represented mathematically as

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

where $\mathbf{x} \in \Re^M$, $\mathbf{s} \in \Re^N$, and $\mathbf{A}$ is an $M \times N$ mixing matrix, find $\mathbf{s}$. Here, $\Re$ denotes the field of real numbers. The class of algorithms that handle such a problem is also called independent component analysis (ICA). When the number of the mixtures is equal to that of the sources (i.e. $M = N$), the objective can be refined to find an $N \times N$ invertible square matrix $\mathbf{W}$ such that

$$\hat{\mathbf{s}} = \mathbf{W}\mathbf{x} \tag{2}$$

where the components of $\hat{\mathbf{s}} = [\hat{s}_1, ..., \hat{s}_N]^T$ are mutually independent as much as possible. This must be done as accurately as possible with the assumption that no more than one source has a Gaussian distribution. Current algorithms can meet this objective within a permutation and scaling of the original sources.

When the number of the sources is larger than that of mixtures, in other words, the mixing matrix $\mathbf{A}$ is an $M \times N$ matrix with $M < N$, the BSS problem is classified as underdetermined. This problem is generally more difficult to tackle than the complete BSS problem where the number of sources is equal to that of mixtures. In the underdetermined case, the sources can not be found easily, but must be inferred even with the knowledge of the mixing matrix. Many of conventional underdetermined BSS algorithms were proposed, and most were based on a maximum a posteriori probability (MAP). Lewicki et al. [1] developed learning overcomplete representations by applying this approach for the sparse sources that have super-Gaussian distributions. Lee et al. [2] later proposed an algorithm based on the learning overcomplete representation and reported good performance for separating speech sources. This algorithm, however, is ineffective for non-sparse sources. When sources are not sparsely distributed in time domain, algorithms for achieving the sparsity in another domain, such as by wavelet packet transform [3] or by short-time Fourier transform [4], [5], were proposed; however, it is difficult to achieve the sparsity for some sources.

In this paper, a novel method for separating the underdetermined mixtures of sources with both sub- and super-Gaussian distributions is proposed. In order to separate the sub- and super-Gaussian sources in the underdetermined case, generalized Gaussian distribution (GGD) is used to model the source distribution. Since the proposed algorithm is also based on MAP and the GGD, the inferred sources minimize the sum of the absolute value of each source s_i taken the p_ith power for $i = 1, ..., N$, where p_i is a positive parameter that controls the exponential decay rate of GGD for s_i. The value of the parameter p_i is determined by the sign of the kurtosis of the ith inferred source. The proposed algorithm operates iteratively; the sources are inferred based on both an estimated GGD and a mixing matrix given the mixtures, and then the parametric GGD and the mixing matrix are updated using the inferred sources. This procedure is repeated at every iteration.

This paper is organized as follows. Section 2 presents a conventional l_1-norm minimization algorithm for sparse sources. Section 3 presents the GGD. Section 4 presents the proposed source inference algorithm by achieving a minimum $L_{\mathbf{p}}$-norm solution that is newly defined in this paper. Section 5 discusses the proposed underdetermined BSS algorithm. Section 6 shows the simulation results, and Section 7 concludes the paper.

2. MINIMUM l_2- AND l_1-NORM SOLUTIONS

Even when the mixing matrix **A** is known in the underdetermined case, the source vector **s** can not be found directly, but must be inferred non-linearly, since the inverse of **A** does not exist. There are infinitely many solutions for **s**. The unique generalized inverse $\mathbf{A}^{+}$, which is called the Moore-Penrose pseudo inverse matrix, gives a minimum l_2-norm solution. For $M \times N$ matrix **A** with $M < N$, it is given as

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

However, the minimum l_2-norm solution does not produce satisfactory performance in separating the sources, since the estimated sources are estimated as linear combinations of the underdetermined mixtures.

In many underdetermined BSS algorithms, the source vector is inferred by maximizing a posteriori probability of **s**, which is given as

$$\begin{aligned}\hat{\mathbf{s}} &= \arg\max_{\mathbf{s}} P(\mathbf{s}|\mathbf{x},\mathbf{A}) \\ &= \arg\max_{\mathbf{s}} P(\mathbf{x}|\mathbf{A},\mathbf{s})P(\mathbf{s})\end{aligned} \tag{4}$$

where $\hat{\mathbf{s}}$ is the most probable decomposition of the mixture. When noise is absent in the mixing process as represented in (1), the maximizing procedure of (4) can be rewritten as

$$\hat{\mathbf{s}} = \arg\max_{\mathbf{s}} P(\mathbf{s}) \tag{5}$$

subject to $\mathbf{As} = \mathbf{x}$. When $P(\mathbf{s})$ is Gaussian, the solution is given as

$$\hat{\mathbf{s}} = \arg\min_{\mathbf{s}} ||\mathbf{s}||_2 \tag{6}$$

subject to $\mathbf{As} = \mathbf{x}$ where the notation $||\cdot||_q$ denotes the l_q-norm, defined as $||\mathbf{s}||_q = (\sum_{i=1}^{N} |s_i|^q)^{1/q}$. This solution is also derived from (3) and likewise does not produce satisfactory performance, since all the sources are assumed Gaussian.

In conventional approach that assumes that the sources are distributed sparsely, it is assumed that the sources follow independently identically distributed (i.i.d.) Laplace distribution [1], [6] which is represented as

$$\begin{aligned}P(\mathbf{s}) &= \prod_{i=1}^{N} \frac{1}{2\alpha} \exp\left\{-\frac{|s_i|}{\alpha}\right\} \\ &= \frac{1}{2^N \alpha^N} \exp\left\{-\frac{1}{\alpha}\sum_{i=1}^{N}|s_i|\right\}\end{aligned} \tag{7}$$

where α is a variance parameter. Using this prior distribution leads to the minimum l_1-norm estimates of the sources as

$$\begin{aligned}\hat{\mathbf{s}} &= \arg\max_{\mathbf{s}} \frac{1}{2^N \alpha^N} \exp\left\{-\frac{1}{\alpha}\sum_{i=1}^{N}|s_i|\right\} \\ &= \arg\min_{\mathbf{s}} ||\mathbf{s}||_1\end{aligned} \tag{8}$$

subject to $\mathbf{As} = \mathbf{x}$ [1], [2], [4], [6], [7]. Unlike the Gaussian prior, the minimum l_1-norm solution can not be obtained by a simple linear operation, and it is well known that the solution can be obtained by a linear programming [8].

3. GENERALIZED GAUSSIAN DISTRIBUTION

In order to unmix the mixtures that are composed of sub- and super-Gaussian sources, a probability density function (*pdf*) that can model both sub- and super-Gaussian distribution is utilized. In this paper, the source distribution is modeled by GGD in [9]. It can generate various symmetric unimodal sub- and super-Gaussian distributions by adjusting the parameters of p and r in

$$P(s;p,r) = \frac{rp}{2\Gamma(1/p)} \exp\left\{-(r|s|)^p\right\} \tag{9}$$

where $\Gamma(\cdot)$ is the Gamma function, which is given as $\Gamma(x) = \int_0^\infty t^{x-1}\exp(-t)dt$. In GGD, p is a positive parameter that controls the distribution's exponential rate of decay, and r is a scaling function of the variance σ^2 of the distribution and p, which is defined as

$$r \equiv \sigma^{-1}\left(\frac{\Gamma(3/p)}{\Gamma(1/p)}\right)^{1/2}. \tag{10}$$

For simplicity, r is set to one in this work. The parameter p controls the distance to the nearest normal distribution. Fig. 1 shows the shape of the GGD according to the value of p. As shown, when p=2, the distribution is a Gaussian normal distribution; when p=1, the Laplacian distribution

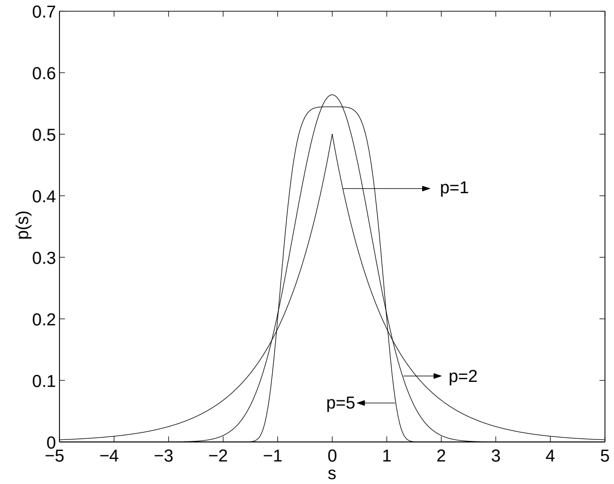

Fig. 1. The shapes of the GGD according to the values of p with r=1.

is obtained; and when p tends to infinity, the distribution becomes uniform.

In contrast to the cases of minimum l_1- and l_2-norm solutions, each source is allowed to have different distributions by the parameter p_i. Therefore, the joint prior distribution of the source random vector **s** is given as

$$P(\mathbf{s}) = \prod_{i=1}^{N} \frac{p_i}{2\Gamma(1/p_i)} \exp\left\{-|s_i|^{p_i}\right\} \tag{11}$$

$$= \left(\frac{1}{2^N}\prod_{j=1}^{N}\frac{p_j}{\Gamma(1/p_j)}\right)\left(\exp\left\{-\sum_{i=1}^{N}|s_i|^{p_i}\right\}\right) \tag{12}$$

with an assumption $r = 1$. The components of the sources are assumed to be mutually independent.

4. MINIMUM $L_\mathbf{p}$-NORM SOLUTION

Using the simplified GGD with $r = 1$ in the previous section, the maximizing procedure of (5) can be represented as

$$\hat{\mathbf{s}} = \arg\max_{\mathbf{s}} \left(\frac{1}{2^N}\prod_{j=1}^{N}\frac{p_j}{\Gamma(1/p_j)}\right)\left(\exp\left\{-\sum_{i=1}^{N}|s_i|^{p_i}\right\}\right) \tag{13}$$

$$= \arg\min_{\mathbf{s}} \sum_{i=1}^{N}|s_i|^{p_i} \tag{14}$$

subject to $\mathbf{As} = \mathbf{x}$ where $\mathbf{s} = [s_1, ..., s_N]^T$. Let the N-dimensional vector be $\mathbf{p} = [p_1, ..., p_N]^T$. In this paper, we introduce a new $L_\mathbf{p}$-norm defined as follows

$$||\mathbf{s}||^\mathbf{p} \equiv \sum_{i=1}^{N}|s_i|^{p_i}. \tag{15}$$

The proposed algorithm leads to source estimates that minimize the $L_\mathbf{p}$-norm which is given as follows

$$\hat{\mathbf{s}} = \arg\min_{\mathbf{s}} ||\mathbf{s}||^\mathbf{p} \tag{16}$$

subject to $\mathbf{As} = \mathbf{x}$. When $\mathbf{p} = [2, ..., 2]^T$, the solution of (14) gives a minimum l_2-norm solution; and when $\mathbf{p} = [1, ..., 1]^T$, a minimum l_1-norm solution is obtained. The minimum $L_\mathbf{p}$-norm solution was achieved using non-linear optimization subroutine in MATLAB.

The newly defined $L_\mathbf{p}$-norm can be rewritten as follows

$$||\mathbf{s}||^\mathbf{p} = \sum_{i=1}^{N}|s_i|^{p_i-q}|s_i|^q$$

$$= \sum_{i=1}^{N} w_i|s_i|^q \tag{17}$$

where $w_i = |s_i|^{p_i-q}$. Therefore, $L_\mathbf{p}$-norm can be represented as the weighted l_q-norm. If the weights are given as $w_i^* = |s_i^*|^{p_i-q}$ where s_i^* is an optimal solution of $L_\mathbf{p}$-norm, the minimum $L_\mathbf{p}$-norm solution can be considered as a minimum weighted l_q-norm solution. This solution can not be found in one step, since s_i^* is required to compute the appropriate weights. In iterative reweighted l_q-norm minimization procedure, the current solution is used to compute a weight which is used for the next iteration.

5. PROPOSED UNDERDETERMINED BSS ALGORITHM

In order to separate the underdetermined mixtures of the sources with both sub- and super-Gaussian distributions, a novel algorithm based on the GGD is proposed. The proposed algorithm is an iterative method of inferring the sources and updating the parameter of the source *pdf* and the mixing matrix. The proposed underdetermined BSS algorithm infers the sources to minimize the $L_\mathbf{p}$-norm of the sources. The performance of the proposed algorithm depends on the accuracy of estimated mixing matrix and p_i.

First, p_i is set to 3 for sub-Gaussian source and 1 for super-Gaussian source. Whether the source s follows sub-Gaussian or super-Gaussian distribution is determined by the sign of the normalized kurtosis, which is defined as

$$\kappa(s) = \frac{E\{s^4\}}{(E\{s^2\})^2} - 3 \tag{18}$$

when s has zero mean. When $\kappa(s)$ has a positive sign, a negative sign, and zero value, s is a super-Gaussian, sub-Gaussian, and Gaussian signal, respectively.

Second, the mixing matrix **A** is updated by the learning rule. It is obtained by differentiating $\log P(\mathbf{x}|\mathbf{A})$ with respect to **A**, which leads to the follows [1]:

$$\begin{aligned} \Delta\mathbf{A} &= \mathbf{A}\mathbf{A}^T \nabla \log\ P(\mathbf{x}|\mathbf{A}) \\ &\approx -\mathbf{A}(\varphi(\hat{\mathbf{s}})\hat{\mathbf{s}}^T + \mathbf{I}) \end{aligned} \tag{19}$$

where **I** represents the identity matrix and the nonlinear function $\varphi(\hat{\mathbf{s}}) = [\varphi(\hat{s}_1), ..., \varphi(\hat{s}_N)]^T$ is given by

$$\varphi(\hat{s}_i) = \frac{\partial \log p_s(\hat{s}_i)}{\partial \hat{s}_i} \tag{20}$$

$$= -p_i|\hat{s}_i|^{p_i-1}\mathrm{sign}(\hat{s}_i) \tag{21}$$

for $i = 1, ..., N$. This learning rule follows the natural gradient proposed by Amari [10] to contain no matrix inverses and achieve a good convergence rate.

The proposed underdetermined BSS algorithm using GGD is summarized as follows. First, the parameter p_i and the mixing matrix **A** are initialized. Then, given the mixtures **x**, the sources are inferred to minimize the $L_\mathbf{p}$-norm of the sources **s** subject to **As** = **x**. The parameter p_i of GGD is determined by the sign of the kurtosis of the inferred source $\hat{\mathbf{s}}$, and the mixing matrix **A** is updated following the learning rule of (19). The procedure is repeated at every iteration.

6. SIMULATIONS

In this section, simulation results are shown to verify the performance of the proposed underdetermined BSS algorithm. The proposed algorithm was compared to the l_1-norm minimization algorithm based on the learning overcomplete representations [2].

First, separation of 2 × N underdetermined mixtures of only sub-Gaussian sources for N = 3, 4, 5, and 6 was performed using the proposed algorithm and the l_1-norm minimization algorithm. The sub-Gaussian sources were generated by taking an inverse hyperbolic sine, $\sinh^{-1}(s_g)$, from a normally distributed random variable s_g. The data of length 5000 was used in the learning process. The batch size was 500. The estimated mixing matrix and the parameter of GGD were updated every batch. In Table 1, the mixing matrix **A** which were used for the simulations for N = 3, 4, 5, and 6 are given.

The column vectors of the mixing matrix to be estimated were initialized as random normalized vectors with the constraint that no two vectors be closer than 30 degrees. The performance of the proposed method was evaluated in terms of signal-to-interference ratio (SIR) for each source. Given

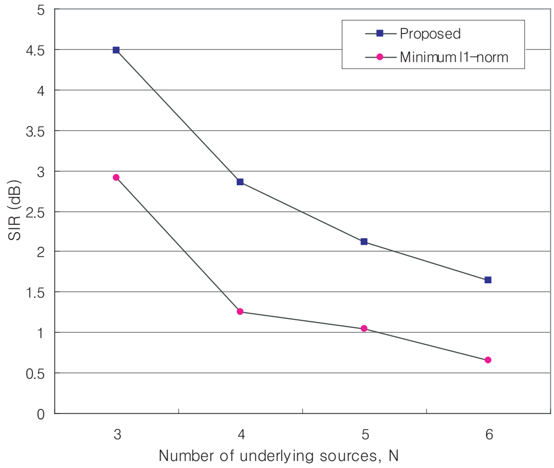

Fig. 2. The performances of the proposed algorithm and the l_1-norm minimization algorithm to separate the underdetermined mixtures of the sub-Gaussian sources. The averaged SIRs according to the number of underlying sources are plotted.

an original source s_j and its estimate $\hat{s}_j$, SIR in dB is defined as

$$\mathrm{SIR} = 10\log_{10}\frac{\sum_{j=1}^{N} E\{s_j^2\}}{\sum_{j=1}^{N} E\{(s_j - \hat{s}_j)^2\}}. \tag{22}$$

Fig. 2 shows the averaged SIRs for ten data according to the number of the underlying sources for the proposed algorithm and the l_1-norm minimization algorithm. The proposed algorithm performed better than the l_1-norm minimization algorithm in separating the underdetermined mixtures of the sub-Gaussian sources. As the number of the sources increased, the separation performance also degraded.

Next, 2×3 underdetermined mixtures of the sources that had different distributions were separated. In generating super-Gaussian sources, a hyperbolic sine, $\sinh(s_g)$, was used instead of $\sinh^{-1}(s_g)$. Two sub- and one super-Gaussian sources were artificially generated. In Table 2, the simulation results are summarized; κ_i^o, κ_i^p, and κ_i^1 represents the kurtosis of the i^{th} original source, estimated source using the proposed algorithm, and estimated source using the l_1-norm minimization algorithm, respectively. SIR_{Lp} and SIR_{l1} denotes SIRs using the proposed algorithm and the l_1-norm minimization algorithm, respectively.

Table 2 shows that only the proposed algorithm can separate the sub-Gaussian sources from the underdetermined mixtures, and the l_1-norm minimization algorithm can not separate the sub-Gaussian sources. The proposed algorithm achieved SIR of 4.844 dB, and l_1-norm minimization algo-

Table 1. The mixing matrix **A** that was used in the simulations for N = 3, 4, 5, and 6.

N	**A**
3	$\begin{bmatrix} \cos(\pi/4) & \cos(\pi/2) & \cos(3\pi/4) \\ \sin(\pi/4) & \sin(\pi/2) & \sin(3\pi/4) \end{bmatrix}$
4	$\begin{bmatrix} \cos(\pi/6) & \cos(\pi/3) & \cos(2\pi/3) & \cos(5\pi/6) \\ \sin(\pi/6) & \sin(\pi/3) & \sin(2\pi/3) & \sin(5\pi/6) \end{bmatrix}$
5	$\begin{bmatrix} \cos(\pi/6) & \cos(\pi/3) & \cos(\pi/2) & \cos(2\pi/3) & \cos(5\pi/6) \\ \sin(\pi/6) & \sin(\pi/3) & \sin(\pi/2) & \sin(2\pi/3) & \sin(5\pi/6) \end{bmatrix}$
6	$\begin{bmatrix} \cos(0) & \cos(\pi/6) & \cos(\pi/3) & \cos(\pi/2) & \cos(2\pi/3) & \cos(5\pi/6) \\ \sin(0) & \sin(\pi/6) & \sin(\pi/3) & \sin(\pi/2) & \sin(2\pi/3) & \sin(5\pi/6) \end{bmatrix}$

Table 2. Performance comparison among the proposed algorithm, and l_1-norm minimization algorithm.

i	κ_i^o	κ_i^p	κ_i^1	SIR_{Lp}	SIR_{l1}
1	-0.826	-0.641	3.907		
2	-0.904	-0.607	10.55	4.844	3.784
3	21.45	11.92	16.17		

rithm SIR of 3.784 dB.

Finally, the separation of three speech signals from two mixtures was performed using the proposed $L_\mathbf{p}$-norm minimization and the l_1-norm minimization algorithm. The length of the speech signals that were sampled at 8 kHz was 10000. The mixing matrix in Table 1 for N=3 was used. In this simulation, the switching condition of the proposed algorithm for determining the parameter **p** was changed. For the source that has relatively high sparsity, the parameter p was set to 0.8, and otherwise 1.0; if the normalized kurtosis value of the inferred source was larger than 2, the parameter was set to 0.8. Fig. 3 shows the three original speech signals and three inferred speech signals using both algorithms after reordering and scaling.

Three speech signals were separated from the two mixtures using both algorithms. The performance of the proposed algorithm was about 1 dB better than l_1-norm minimization algorithm: the proposed achieved SIR of 4.012 dB, and l_1-norm minimization algorithm SIR of 3.057 dB.

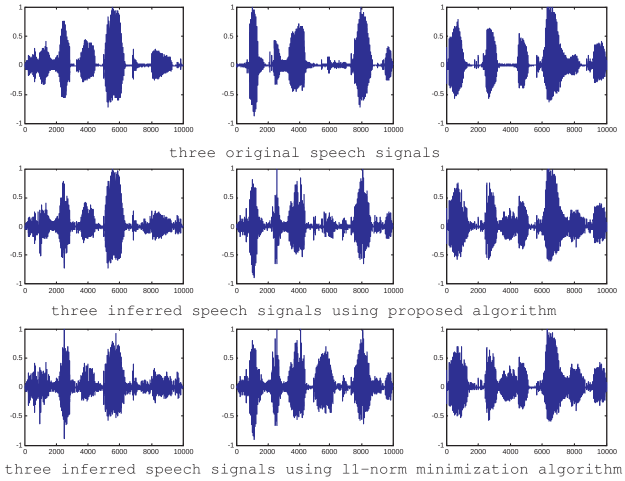

Fig. 3. The separation of three speech signals from two mixtures. Top row: three original speech signals. Middle row: three inferred speech signals using the proposed algorithm. Bottom row: three inferred speech signals using the l_1-norm minimization algorithm.

7. CONCLUSIONS

A novel method for separating underlying sources with both sub- and super-Gaussian distributions from the underdetermined mixtures is proposed. The generalized Gaussian distribution (GGD) is used to model simultaneously both sub- and super-Gaussian distributions. The process of finding the most probable decomposition of the mixtures based on the GGD leads to that of minimizing the $L_\mathbf{p}$-norm of the estimated sources that was defined newly in this paper. The switching condition for determining the decay rate of the GGD is made according to the sign of the kurtosis of the inferred source. Using the proposed algorithm, sources were separated from the underdetermined mixtures that were composed of all the sub-Gaussian sources and both sub- and super-Gaussian sources. However, the l_1-minimization algorithm failed to separate the sub-Gaussian sources from those mixtures. In separation of three speech sources from two mixtures, the proposed algorithm applied different p_i value according to the normalized kurtosis value of the inferred source $\hat{s}_i$, and achieved about 1 dB better SIR than the l_1-norm minimization algorithm.

The parameter **p** to adjust the decay rate of the GGD was determined by the sign of the kurtosis of the inferred source. In order to obtain a flexible model for the source distribution, an update routine for the source distribution is involved. The proposed algorithm was compared to only the

l_1-norm minimization algorithm and should be compared to other underdetermined BSS algorithms in the future.

8. ACKNOWLEDGMENTS

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.

9. REFERENCES

[1] M. S. Lewicki and T. J. Sejnowski, "Learning overcomplete representations," *Neural Computation*, vol. 12, pp. 337–365, 2000.

[2] 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 Processing Letters*, vol. 6, no. 4, pp. 87–90, 1999.

[3] Y. Li, A. Cichocki, and S. I. Amari, "Sparse component analysis for blind source separation with less sensors than sources," in *4th International Symposium on Independent Component Analysis and Blind Signal Separation*, 2003, pp. 89–94.

[4] P. Bofill and M. Zibulevsky, "Underdertermined blind source separation using sparse representations," *Signal Processing*, vol. 81, pp. 2353–2362, 2001.

[5] O. Yilmaz and S. Rickard, "Blind separation of speech mixtures via time-frequency masking," *IEEE Trans. Signal Processing*, vol. 52, no. 7, pp. 1830–1847, 2004.

[6] I. Takigawa, M. Kudo, and J. Toyama, "Performance analysis of minimum l_1-norm solutions for underdetermined source separation," *IEEE Trans. Signal Processing*, vol. 52, no. 3, pp. 582–591, 2004.

[7] F. J. Theis, E. W. Lang, and C. G. Puntonet, "A geometric algorithm for overcomplete linear ICA," *Neurocomputing*, vol. 56, pp. 381–398, 2004.

[8] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *Soceity for Industrial and Applied Mathematics*, vol. 20, no. 1, pp. 33–61, 1998.

[9] J. H. Miller and J. B. Thomas, "Detectors for discrete-time signals in non-Gaussian noise," *IEEE Trans. Information Theory*, vol. IT-18, no. 2, pp. 241–250, 1972.

[10] S. I. Amari, "Natural gradient works efficiently in learning," *Neural computation*, vol. 10, pp. 251–276, 1998.