Equivalence Probability and Sparsity of Two Sparse Solutions in Sparse Representation

Yuanqing Li, *Member, IEEE*, Andrzej Cichocki, *Senior Member, IEEE*, Shun-ichi Amari, *Life Fellow, IEEE*, Shengli Xie, *Senior Member, IEEE*, and Cuntai Guan, *Senior Member, IEEE*

***Abstract*—This paper discusses the estimation and numerical calculation of the probability that the 0-norm and 1-norm solutions of underdetermined linear equations are equivalent in the case of sparse representation. First, we define the sparsity degree of a signal. Two equivalence probability estimates are obtained when the entries of the 0-norm solution have different sparsity degrees. One is for the case in which the basis matrix is given or estimated, and the other is for the case in which the basis matrix is random. However, the computational burden to calculate these probabilities increases exponentially as the number of columns of the basis matrix increases. This computational complexity problem can be avoided through a sampling method. Next, we analyze the sparsity degree of mixtures and establish the relationship between the equivalence probability and the sparsity degree of the mixtures. This relationship can be used to analyze the performance of blind source separation (BSS). Furthermore, we extend the equivalence probability estimates to the small noise case. Finally, we illustrate how to use these theoretical results to guarantee a satisfactory performance in underdetermined BSS.**

***Index Terms*—Blind source separation (BSS), equivalence probability, noise, sparse representation, sparsity degree, 0-norm solution, 1-norm solution.**

I. INTRODUCTION

In recent years, there have been many studies on the sparse representation (or sparse component analysis, sparse coding) of signals (e.g., [1]–[10]). Sparse representation of signals can be modeled by

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

where $\mathbf{x}$ is a given n-dimensional signal vector, $\mathbf{A} \in \mathbf{R}^{n\times m}$ $(n < m)$ is a basis matrix with full row rank, and $\mathbf{s} \in \mathbf{R}^m$ is a coefficient vector.

Among the infinite possible solutions (1), the sparsest one, which is the solution of the following optimization problem, has received much attention

$$(P_0) \quad \min \|\mathbf{s}\|_0, \text{ s.t. } \mathbf{A}\mathbf{s} = \mathbf{x}$$

where 0-norm $\|\mathbf{s}\|_0$ is the number of nonzero elements of $\mathbf{s}$. It is well known that the optimization problem (P_0) is NP-hard, i.e., the computational time (or the computational burden) for solving this optimization problem increases exponentially as m increases (in other words, the computational time is nonpolynomial).

In many references, e.g., [1], a basis pursuit (BP) algorithm was presented by solving the following optimization problem, which can be transformed into a linear programming problem:

$$(P_1) \quad \min \|\mathbf{s}\|_1 = \sum_{i=1}^{m} |s_i|, \text{ s.t. } \mathbf{A}\mathbf{s} = \mathbf{x}.$$

Hereafter, the solutions of (P_0) and (P_1) are called the 0-norm solution and 1-norm solution, respectively.

Many studies have shown that in many cases the BP algorithm can also find the sparsest solution, i.e., the 0-norm solution. By studying the equivalence between the 0-norm solution and the 1-norm solution, several conditions have been established that guarantee the BP's success (e.g., [5] and [10]). These conditions, however, are often too strict to reflect the fact that the BP algorithm often finds the 0-norm solution. Recently, researchers discussed the equivalence between the 0-norm and 1-norm solutions within a probabilistic framework (e.g., [11] and [15]–[17]). The condition, under which the two sparse solutions are equivalent with a high probability, is much weaker than those obtained in the deterministic case. In this paper, the probability that the 0-norm solution and the 1-norm solution are equivalent is called the equivalence probability. In blind source separation (BSS), this probability is also called the recoverability probability since the 0-norm solution is a source vector in most cases. As will be seen, the equivalence probability (or recoverability probability) has different formula depending on various conditions.

An important application of the sparse representation is in underdetermined BSS, in which the number of sources (m) is greater than the number of observations (n). Note that (1) can be taken as a mixing model, in which $\mathbf{x}$ is a mixture signal vector, $\mathbf{A}$ is an unknown mixing matrix, and $\mathbf{s}$ is an unknown source vector. If a source vector $\mathbf{s}$ is sufficiently sparse in the analyzed

Manuscript received December 15, 2007; revised May 05, 2008; accepted June 03, 2008. Current version published November 28, 2008. The work of Y. Li was supported in part by the National Natural Science Foundation of China under Grants 60475004 and 60825306, by the Program for New Century Excellent Talents in China University under Grant NCET-04-0816, by Guangdong Province Science Foundation for Research Team Program under Grant 04205783.

Y. Li was with the Institute for Infocomm Research, Singapore 119613, Singapore and RIKEN Brain Science Institute, Saitama 3510198, Japan. He is now with South China University of Technology, Guangzhou 510640, China (e-mail: auyqli@scut.edu.cn).

A. Cichocki is with Laboratory for Advanced Brain Signal Processing, RIKEN Brain Science Institute, Wako shi, Saitama 3510198, Japan (e-mail: cia@brain.riken.jp).

S. Amari is with the Unit for Mathematical Neuroscience, RIKEN Brain Science Institute, Wako shi, Saitama 3510198, Japan (e-mail: amari@brain.riken.jp).

S. Xie is with South China University of Technology, Guangzhou 510640, China (e-mail: adshlxie@scut.edu.cn).

C. Guan is with Infocomm Research, Agency for Science, Technology and Research (A*STAR), Singapore 119613, Singapore (e-mail: ctguan@i2r.a-star.edu.sg).

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/TNN.2008.2003980

domain, then it is the 0-norm solution. The equivalence problem of the 0-norm and 1-norm solutions then becomes a recoverability problem here. One can first estimate the mixing matrix and then recover the source vector using the BP algorithm [11], [16], [19].

In [16] and [17], underdetermined BSS is discussed using a two-stage sparse representation approach. In [16], an algorithm for estimating the mixing matrix was first developed. A necessary and sufficient condition for recoverability of a source vector was then obtained. This condition shows that the recoverability of the source vector (i.e., the 0-norm solution in this paper) only depends on its sign pattern. Based on this condition and various types of source sparsity, several probability inequalities and probability estimates for the recoverability were established when the mixing matrix $\mathbf{A}$ is given or estimated. A general case where the mixing matrix is random was discussed in [17]. A new necessary and sufficient condition for recoverability of a source vector was first presented. Using this condition, several theoretical probability estimates on recoverability were obtained. The obtained recoverability probability estimates include $P(n, m, l; \mathbf{A})$, $P(n, m, \alpha; \mathbf{A})$, $P(n, m, \lambda; \mathbf{A})$ [16], $P(n, m, l)$, $P(n, m, \alpha)$, and $P(n, m, \lambda)$ [17], where n is the number of the mixtures, m is the number of sources, l is the number of nonzeros of a random source vector, α is the probability that each entry of the source vector is equal to zero, and λ is a parameter of a Laplacian distribution from which the entries of the source vector are drawn. There still remain several problems in [16] and [17], which require extensive discussions. These problems include the following.

1) Computational complexity: the computational burden for calculating those above probability estimates grows exponentially with increasing m.
2) The comparison of these probability estimates: through this comparison, we may find one that reflects the reality.
3) Noisy case.
4) Applications of these probability estimates in BSS (to guarantee the performance of BSS).
5) A general case in which the sources are from different distributions (this means that the degrees of sparsity are different depending on the sources).

This paper focuses on the above problems. The recoverability probability of a source vector $\mathbf{S}^{(0)}$ depends on l, the number of its nonzeros. Therefore, the probabilities $P(n, m, l; \mathbf{A})$ and $P(n, m, l)$ play an important role in [16] and [17], which reflect the recoverability of a single source vector. Since l may be different for different source vectors, $P(n, m, l)$ [or $P(n, m, l; \mathbf{A})$] cannot reflect the recoverability of the sources in a time interval (each time point corresponds to a source vector). The probability estimate $P(n, m, \lambda)$ [or $P(n, m, \lambda; \mathbf{A})$] holds under the assumption that all sources are drawn from a standard Laplacian distribution. This assumption may not be strictly true in many cases. As will be seen, the probability estimate $P(n, m, \alpha)$ [or $P(n, m, \alpha; \mathbf{A})$] can reflect the recoverability of sources in a time interval. Furthermore, it is a convenient method to analyze the performance of BSS.

The remainder of this paper is organized as follows. In Section II, a sparsity degree of a random variable is defined. When all entries of a 0-norm solution (seen as a source vector) have different sparsity degrees, the recoverability probability estimate is then established. This probability estimate reflects the relationship between the equivalence probability and the sparsity degree vector of the 0-norm solution (i.e., the source vector in BSS). When m is large, a sampling method is proposed to obtain this probability estimate. In Section III, another relationship between the equivalence probability and the sparsity degree of observed mixtures is established. These two relationships can be used to analyze the performance of BSS. In Section IV, the results on the probability estimates and the two relationships are extended to the case when the noise is small, i.e., small noise case. Section V shows how to use these results to guarantee the performance of sparse representation-based BSS.

II. Sparsity Degree and Equivalence

In this section, we first define the sparsity degree of a signal and establish a relationship between the sparsity degree of the 0-norm solution and the equivalence probability.

Discussion in the following sections are based on the following optimization problem:

$$(P_1') \quad \min \sum_{i=1}^{m} |s_i|, \text{ s.t. } \mathbf{A}\mathbf{s} = \mathbf{x}^*$$

where $\mathbf{x}^* = \mathbf{A}\mathbf{s}^{(0)}$, $\mathbf{s}^{(0)} \in \mathbf{R}^m$, is the 0-norm solution corresponding to $\mathbf{x}^*$. Here $\mathbf{x}^*$ is seen as generated by the vector $\mathbf{s}^{(0)}$ (0-norm solution). The solution of (P_1') is denoted as $\mathbf{s}^{(1)}$. If $\mathbf{s}^{(1)} = \mathbf{s}^{(0)}$, then we say that $\mathbf{s}^{(0)}$ is recovered by solving (P_1').

Definition 1:

1) For a random variable z, the probability $P(z = 0)$ is called the sparsity degree of z.
2) For a signal $\mathbf{y} = (y(1), y(2), \ldots)$ (a sample sequence), suppose that the probability $P(y(k) = 0)$ is the same for all k. Then, the probability $P(y(k) = 0)$ is called the sparsity degree of $\mathbf{y}$.

Remark 1:

i) In many real-world applications, e.g., where noise exists, the probability that $z = 0$ or $y(k) = 0$ is close to or equal to zero; however, z or $\mathbf{y}$ is still sparse since $|z|$ or $|y(k)|$ takes on small values with a high probability. In this case, Definition 1 is modified as follows. A small positive constant ϵ is first determined according to the specific task, and the probability $P(|z| < \epsilon)$ or $P(|y(k)| < \epsilon)$ is taken as sparsity degree of z or $\mathbf{y}$. Example 3 in Section V is an illustration of this case.
ii) For a given signal $\mathbf{y}$, its sparsity degree can be estimated using the ratio of the number of its zeros to the total number of sample points, i.e., the frequency of zeros in $\mathbf{y}$.

For a random 0-norm solution $\mathbf{s}^{(0)}$ in (P_1'), suppose that the sparsity degree of its ith entry $s_i^{(0)}$ is $\alpha_i^{(s)}$, $i = 1, \ldots, m$. Here the m entries of the 0-norm solution $\mathbf{s}^{(0)}$ are taken as m random variables. In BSS, $\alpha_i^{(s)}$ represents the sparsity degree of the ith source ($i = 1, \ldots, m$). Now we discuss the relationship between the recoverability probability of the 0-norm solution $\mathbf{s}^{(0)}$ and its sparsity degree vector $[\alpha_1^{(s)}, \ldots, \alpha_m^{(s)}]$.

Denote the index set by $G = \{1, \ldots, m\}$. Using G, we define m sets $F_j = \{[i_1, \ldots, i_j] | i_1, \ldots, i_j \in G\}$,

$j = 1, \ldots, m$ (each set F_j contains j indices). Obviously, F_j contains C_m^j $(= m!/(m-j)!j!)$ vectors. For a vector $[i_1, \ldots, i_j] \in F_j$, we define a set of sign vectors with j nonzeros as $T_{i_1,\ldots,i_j} = \{\mathbf{t} = [t_1, \ldots, t_m]^T | t_i \in \{-1, 1\}$, if $i \in \{i_1, \ldots, i_j\}$, otherwise $t_i = 0\}$. There are 2^j sign vectors in $T_{i_1,\ldots,i_j}$. Let $T_l = \{\mathbf{t} = [t_1, \ldots, t_m]^T | t_j \in \{-1, 0, 1\}$, $\sum_{j=1}^m |t_j| = l\}$, which is the set of sign vectors with l nonzero entries.

As a preparation, we first quote a result in [16]. For a 0-norm solution $\mathbf{s}^{(0)}$, we denote its sign vector as $\mathrm{sign}(\mathbf{s}^{(0)})$. Then, $\mathbf{s}^{(0)}$ can be recovered by solving (P_1') if and only if $\mathrm{sign}(\mathbf{s}^{(0)})$ can be recovered by solving (P_1'). According to this conclusion, the recoverability of $\mathbf{s}^{(0)}$ depends only on its sign pattern.

The following theorem contains two conclusions that are extensions of the results in [16] and [17], respectively.

Theorem 1:

1) For a given basis matrix $\mathbf{A}$, we have the following equivalence probability estimate:

$$P\left(\mathbf{s}^{(0)} = \mathbf{s}^{(1)}; \alpha_1^{(s)}, \ldots, \alpha_m^{(s)}, \mathbf{A}\right) = \prod_{j=1}^m \alpha_j^{(s)} + \sum_{j=1}^m \sum_{[i_1,\ldots,i_j]\in F_j} \left(1-\alpha_{i_1}^{(s)}\right)\cdots\left(1-\alpha_{i_j}^{(s)}\right) \prod_{k=1,k\neq i_1,\ldots,i_j}^m \alpha_k^{(s)} \frac{q^{(j)}(i_1,\ldots,i_j)}{2^j} \tag{2}$$

where $q^{(j)}(i_1, \ldots, i_j)$ is the number of sign vectors in $T_{i_1,\ldots,i_j}$, which can be recovered by solving (P_1').

2) For random basis matrix $\mathbf{A}$, we have

$$P\left(\mathbf{s}^{(0)} = \mathbf{s}^{(1)}; \alpha_1^{(s)}, \ldots, \alpha_m^{(s)}\right) = \prod_{j=1}^m \alpha_j^{(s)} + \sum_{j=1}^m \sum_{[i_1,\ldots,i_j]\in F_j} \left(1-\alpha_{i_1}^{(s)}\right)\cdots\left(1-\alpha_{i_j}^{(s)}\right) \prod_{k=1,k\neq i_1,\ldots,i_j}^m \alpha_k^{(s)} P(n,m,j) \tag{3}$$

where $P(n, m, j)$ $(j = 1, \ldots, m)$ is the probability that a 0-norm solution with j nonzeros can be recovered by solving (P_1') (the estimation of $P(n, m, j)$ was discussed in [17]. See Appendix I in this paper).

Proof:

1) For any 0-norm solution $\mathbf{s}^{(0)} \in R^m$, we have

$$P\left(s_k^{(0)} = 0\right) = \alpha_k^{(s)} \quad P\left(s_k^{(0)} \neq 0\right) = 1 - \alpha_k^{(s)} \tag{4}$$

where $k = 1, \ldots, m$.

Denote $\mathrm{Non}(\mathbf{s}^{(0)})$ as the set of indices of the nonzero elements in $\mathbf{s}^{(0)}$. Then, the probability that $\mathrm{Non}(\mathbf{s}^{(0)}) = \{i_1, \ldots, i_j\}$ is $(1 - \alpha_{i_1}^{(s)}) \cdots (1 - \alpha_{i_j}^{(s)}) \prod_{k=1,k\neq i_1,\ldots,i_j}^m \alpha_k^{(s)}$. Furthermore, if there are $q^{(j)}(i_1, \ldots, i_j)$ sign vectors in $T_{i_1,\ldots,i_j}$, which can be recovered by solving (P_1'), then the recoverability probability of a sign vector randomly taken in $T_{i_1,\ldots,i_j}$ is $q^{(j)}(i_1, \ldots, i_j)/2^j$, i.e.,

$$P\left(\mathbf{s}^{(0)} = \mathbf{s}^{(1)}; \mathrm{Non}\left(\mathbf{s}^{(0)}\right) = [i_1, \ldots, i_j], \mathbf{A}\right) = \frac{q^{(j)}(i_1, \ldots, i_j)}{2^j}. \tag{5}$$

Thus, we have

$$\begin{aligned}
&P\left(\mathbf{s}^{(0)} = \mathbf{s}^{(1)}; \alpha_1^{(s)}, \ldots, \alpha_m^{(s)}, \mathbf{A}\right)\\
&= \sum_{j=0}^m \sum_{[i_1,\ldots,i_j]\in F_j} P\left(\mathrm{Non}\left(\mathbf{s}^{(0)}\right) = [i_1, \ldots, i_j]\right)\\
&P\left(\mathbf{s}^{(0)} = \mathbf{s}^{(1)}; \mathrm{Non}\left(\mathbf{s}^{(0)}\right) = [i_1, \ldots, i_j], \mathbf{A}\right)\\
&= \prod_{j=1}^m \alpha_j^{(s)} + \sum_{j=1}^m \sum_{[i_1,\ldots,i_j]\in F_j} \left(1-\alpha_{i_1}^{(s)}\right)\cdots\left(1-\alpha_{i_j}^{(s)}\right)\\
&\quad\times \prod_{k=1,k\neq i_1,\ldots,i_j}^m \alpha_k^{(s)} P\left(\mathbf{s}^{(0)} = \mathbf{s}^{(1)};\right.\\
&\qquad\qquad \left.\mathrm{Non}\left(\mathbf{s}^{(0)}\right) = [i_1, \ldots, i_j], \mathbf{A}\right)\\
&= \prod_{j=1}^m \alpha_j^{(s)} + \sum_{j=1}^m \sum_{[i_1,\ldots,i_j]\in F_j} \left(1-\alpha_{i_1}^{(s)}\right)\cdots\left(1-\alpha_{i_j}^{(s)}\right)\\
&\quad\times \prod_{k=1,k\neq i_1,\ldots,i_j}^m \alpha_k^{(s)} \frac{q^{(j)}(i_1,\ldots,i_j)}{2^j}.
\end{aligned} \tag{6}$$

2) Now we consider the case in which the mixing matrix is random. Suppose that $\mathbf{s}^{(0)}$ has j nonzeros with indices $i_1, \ldots, i_j$. Since the mixing matrix is random, the recoverability probability of $\mathbf{s}^{(0)}$ is equal to $P(n, m, j)$, i.e., it does not depend on specific $i_1, \ldots, i_j$. Based on this conclusion and similar derivation as above, we can prove the second part of this theorem. The theorem is thus proved.

Note that the equivalence probability $P(\mathbf{s}^{(0)} = \mathbf{s}^{(1)}; \alpha_1^{(s)}, \ldots, \alpha_m^{(s)}, \mathbf{A})$ is a conditional probability that $\mathbf{s}^{(0)}$ and $\mathbf{s}^{(1)}$ are equivalent under the condition that $\alpha_1^{(s)}, \ldots, \alpha_m^{(s)}$, and $\mathbf{A}$ are given, and so does $P(\mathbf{s}^{(0)} = \mathbf{s}^{(1)}; \alpha_1^{(s)}, \ldots, \alpha_m^{(s)})$.

If $\alpha_1^{(s)} = \alpha_2^{(s)} = \cdots = \alpha_m^{(s)}$ [denoted as $\alpha^{(s)}$], then (2) becomes

$$P\left(\mathbf{s}^{(0)} = \mathbf{s}^{(1)}; \alpha^{(s)}, \mathbf{A}\right) = \sum_{j=0}^m C_m^j \left(1-\alpha^{(s)}\right)^j \left(\alpha^{(s)}\right)^{m-j} \frac{q_j}{C_m^j 2^j} \tag{7}$$

where q_j is the number of sign vectors in the set T_j which can be recovered by solving (P_1').

Furthermore, if $\alpha_1^{(s)} = \alpha_2^{(s)} = \ldots = \alpha_m^{(s)}$ [denoted as $\alpha^{(s)}$], then (3) for a random basis matrix becomes

$$P\left(\mathbf{s}^{(0)} = \mathbf{s}^{(1)}; \alpha^{(s)}\right) = \sum_{j=0}^m C_m^j \left(1-\alpha^{(s)}\right)^j \left(\alpha^{(s)}\right)^{m-j} P(n,m,j). \tag{8}$$

Equations (7) and (8) were obtained in [16] and [17].

When (1) is considered as a mixing model in BSS, then the 0-norm solution can be seen as a source vector. The sparsity degree $\alpha_i^{(s)}$ of the ith entry of the 0-norm solution corresponds to the sparsity degree of the ith source. This sparsity degree can be estimated in the following way. Suppose that there are N_0 time sample points, and that the ith source equals to zero at L_0 sample points. Then, the sparsity degree of the ith source is L_0/N_0. Therefore, the probability estimates in (2) and (3) reflect the recoverability of sources in the corresponding time interval of sampling.

In order to calculate the equivalence probabilities in (2) and (3) [or (7) and (8)], we need to find the number of sign vectors that can be recovered by solving (P_1'). Note that the total number of sign vectors is $\sum_{l=0}^{m} 2^l C_m^l$ for a given m. When m increases, the computational burden for estimating the probabilities in (2) and (3) will increase exponentially. This problem can be resolved by the following sampling method.

Sampling Method:

i) The mixing matrix $\mathbf{A}$ is given. For a sparsity degree vector $[\alpha_1, \ldots, \alpha_m]$, we randomly generate L_0 (e.g., 3000) column vectors $\mathbf{s}^{(k)} \in R^m$ using the uniform distribution in $[-0.5, 0.5]$, $k = 1, \ldots, L_0$. If $|s_i^{(k)}| < 0.5\alpha_i$, we further set $s_i^{(k)} = 0$. Then, the sparsity degree vector of $\mathbf{s}^{(k)}$ is $[\alpha_1, \ldots, \alpha_m]$. Now we check the recoverability of $\mathbf{s}^{(k)}$ by solving (P_1') with $\mathbf{x}^* = \mathbf{A}\mathbf{s}^{(k)}$. Suppose that among the L_0 vectors, there are L_1 vectors recovered. Then, we have an estimate of the probability $P(\mathbf{s}^{(0)} = \mathbf{s}^{(1)}; \alpha_1^{(s)}, \ldots, \alpha_m^{(s)}, \mathbf{A}) \approx (L_1/L_0)$ [denoted as $\mathrm{pro}(\alpha_1, \ldots, \alpha_m, \mathbf{A})$].

ii) The mixing matrix $\mathbf{A}$ is random. Similarly as above, we randomly generate L_0 (e.g., 3000) column vectors $\mathbf{s}^{(k)} \in R^m$ using the uniform distribution in $[-0.5, 0.5]$, $k = 1, \ldots, L_0$ such that the sparsity degree vector of $\mathbf{s}^{(k)}$ is $[\alpha_1, \ldots, \alpha_m]$. Furthermore, we randomly generate L_0 mixing matrices $\mathbf{A}^{(k)}$ using a distribution (e.g., the uniform distribution in $[-0.5, 0.5]$). For each pair $(\mathbf{A}^{(k)}, \mathbf{s}^{(k)})$, we check the recoverability of $\mathbf{s}^{(k)}$ by solving (P_1') with $\mathbf{A}^{(k)}$ and $\mathbf{x}^*$ $(= \mathbf{A}^{(k)}\mathbf{s}^{(k)})$. Suppose that, among the L_0 vectors, there are L_2 vectors recovered. Then, we have an estimation of the probability $P(\mathbf{s}^{(0)} = \mathbf{s}^{(1)}; \alpha_1^{(s)}, \ldots, \alpha_m^{(s)}) \approx (L_2/L_0)$ [denoted as $\mathrm{pro}(\alpha_1, \ldots, \alpha_m)$].

One question arises here: If m increases, should the number of samples (L_0) be increased exponentially? Our theoretical analysis and simulations tell us that it is not necessary to increase L_0 exponentially (see Appendix I).

The validity of the above sampling method will be demonstrated in Example 1.

Equations (2) and (3) reflect the relationship between the equivalence probability and the sparsity degrees of the entries of the 0-norm solutions. Note that the equivalence probabilities in (2) and (3) do not depend on the amplitudes of the nonzero entries of the 0-norm solution. From Theorem 1, we find that the equivalence probabilities in (2) and (3) are determined only by the sparsity degrees of the entries of the 0-norm solution. Therefore, we have the following corollary.

Corollary 1: For a given basis matrix $\mathbf{A}$, the equivalence probability depends only on the sparsity degrees of the entries of the 0-norm solution. Furthermore, for a random basis matrix, the equivalence probability depends on the distribution of basis matrix $\mathbf{A}$ and the sparsity degrees of the entries of the 0-norm solution.

For a random 0-norm solution $\mathbf{s}^{(0)}$ with sparsity degree vector $(\alpha_1^{(s)}, \ldots, \alpha_m^{(s)})$, we calculate two means, the arithmetic mean $\bar{\alpha}^{(s)} = (1/m)\sum_{i=1}^{m} \alpha_i^{(s)}$, and the geometric mean $\tilde{\alpha}^{(s)} = (\prod_{i=1}^{m} \alpha_i^{(s)})^{1m}$.

Since the sparsity degree $\alpha_i^{(s)}$ of a single entry of the 0-norm solution $\mathbf{s}^{(0)}$ is not easy to estimate in real-world applications, we consider the special case in which all entries of the 0-norm solution $\mathbf{s}^{(0)}$ have the same sparsity degree $\bar{\alpha}^{(s)}$ or $\tilde{\alpha}^{(s)}$ (they can be estimated sometimes). Let $\bar{\mathbf{s}}^{(0)}$ and $\tilde{\mathbf{s}}^{(0)}$ denote two random 0-norm solutions where the sparsity degree of each entry of $\bar{\mathbf{s}}^{(0)}$ is $\bar{\alpha}^{(s)}$, while the sparsity degree of each entry of $\tilde{\mathbf{s}}^{(0)}$ is $\tilde{\alpha}^{(s)}$.

It is not difficult to find that the average numbers of zeros of $\mathbf{s}^{(0)}$, $\bar{\mathbf{s}}^{(0)}$, and $\tilde{\mathbf{s}}^{(0)}$ are $\sum_{i=1}^{m} \alpha_i^{(s)}$, $m\bar{\alpha}^{(s)}$ $(= \sum_{i=1}^{m} \alpha_i^{(s)})$, and $m\tilde{\alpha}^{(s)}$ $(= m(\prod_{i=1}^{m} \alpha_i^{(s)})^{1/m})$. Therefore, the mean of the number of the zeros of $\mathbf{s}^{(0)}$ is equal to the mean of the number of the zeros of $\bar{\mathbf{s}}^{(0)}$. Since the arithmetic mean is larger than the geometric mean, the mean of the number of the zeros of $\mathbf{s}^{(0)}$ [or $\bar{\mathbf{s}}^{(0)}$] is larger than the mean of the number of the zeros of $\tilde{\mathbf{s}}^{(0)}$. Based on the analysis here, we have the following approximations and inequalities:

$$P\left(\mathbf{s}^{(0)} = \mathbf{s}^{(1)}; \alpha_1^{(s)}, \ldots, \alpha_m^{(s)}, \mathbf{A}\right) \approx P\left(\bar{\mathbf{s}}^{(0)} = \bar{\mathbf{s}}^{(1)}; \bar{\alpha}^{(s)}, \mathbf{A}\right)$$
$$P\left(\mathbf{s}^{(0)} = \mathbf{s}^{(1)}; \alpha_1^{(s)}, \ldots, \alpha_m^{(s)}\right) \approx P\left(\bar{\mathbf{s}}^{(0)} = \bar{\mathbf{s}}^{(1)}; \bar{\alpha}^{(s)}\right) \quad (9)$$
$$P\left(\bar{\mathbf{s}}^{(0)} = \bar{\mathbf{s}}^{(1)}; \bar{\alpha}^{(s)}, \mathbf{A}\right) > P\left(\tilde{\mathbf{s}}^{(0)} = \tilde{\mathbf{s}}^{(1)}; \tilde{\alpha}^{(s)}, \mathbf{A}\right)$$
$$P\left(\bar{\mathbf{s}}^{(0)} = \bar{\mathbf{s}}^{(1)}; \bar{\alpha}^{(s)}\right) > P\left(\tilde{\mathbf{s}}^{(0)} = \tilde{\mathbf{s}}^{(1)}; \tilde{\alpha}^{(s)}\right) \quad (10)$$

where $\bar{\mathbf{s}}^{(1)}$ and $\tilde{\mathbf{s}}^{(1)}$ are the 1-norm solutions of (P_1') with $\mathbf{x}^* = \mathbf{A}\bar{\mathbf{s}}^{(0)}$ and $\mathbf{x}^* = \mathbf{A}\tilde{\mathbf{s}}^{(0)}$, respectively.

As will be seen in next section, (9) and (10) are convenient to use in real-world applications.

Now we present a simulation example to demonstrate the above results including (9) and (10).

Example 1: In this example, we always set $n = 7$ and $m = 9$. The simulations in this example are in two categories: i) $\mathbf{A}$ is given or estimated and ii) $\mathbf{A}$ is random. In the following, all random matrices $\mathbf{A}$ are generated using a standard normal distribution. They also can be generated using other distributions, e.g., a uniform distribution.

i) First, we randomly generate ten sparsity degree vectors $\alpha^{(j)} = [\alpha_1^{(j)}, \ldots, \alpha_9^{(j)}]$, $j = 1, \ldots, 10$, using the uniform distribution in $[0, 1]$. Calculate the arithmetic mean $\bar{\alpha}^{(j)}$ and the geometric mean $\tilde{\alpha}^{(j)}$ for each $\alpha^{(j)}$.

Second, we randomly give a matrix $\mathbf{A} \in R^{7\times 9}$ of which each entry is generated using a standard normal distribution. Using this matrix and the sparsity degree vectors $\alpha^{(j)}$, we calculate the recoverability probabilities $P(\mathbf{s}^{(0)} = \mathbf{s}^{(1)}; \alpha_1^{(j)}, \ldots, \alpha_9^{(j)}, \mathbf{A})$ according to (2), where $j = 1, \ldots, 10$.

Third, for each $\alpha^{(j)}$ $(j = 1, \ldots, 10)$, we estimate the recoverability probability using the sampling method and obtain $\mathrm{pro}(j; \alpha_1^{(j)}, \ldots, \alpha_9^{(j)}, \mathbf{A})$. In this example, we always arbitrarily set $L_0 = 3000$ for the sampling method.

For each $\bar{\alpha}^{(j)}$ $(j = 1, \ldots, 10)$, we also estimate the recoverability probability using sampling method. Note that the sparsity degree of each entry of the source vector is $\bar{\alpha}^{(j)}$. The obtained probability is denoted as $\bar{\mathrm{pro}}(j; \bar{\alpha}^{(j)}, \mathbf{A})$. Using $\tilde{\alpha}^{(j)}$ to replace $\bar{\alpha}^{(j)}$ $(j = 1, \ldots, 10)$, we perform similar simulation and obtain another probability estimate $\bar{\mathrm{pro}}(j; \tilde{\alpha}^{(j)}, \mathbf{A})$.

The above simulation results are shown in the first row of Fig. 1. The first subplot shows $\bar{\alpha}^{(j)}$ (denoted by "o")

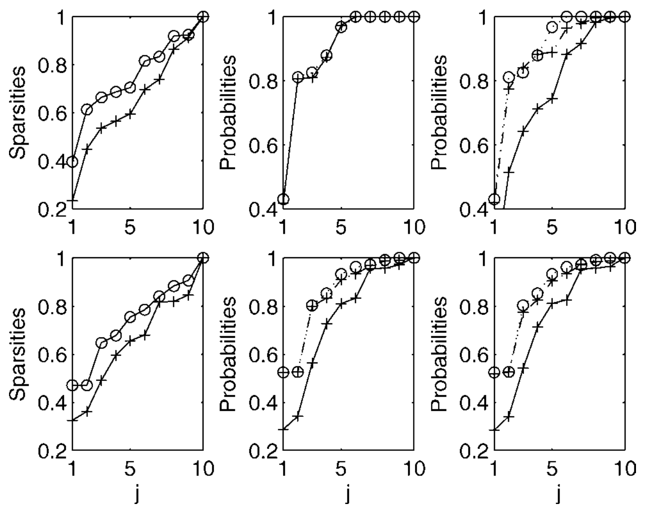

Fig. 1. Simulation results in Example 1. The first row is for the case, when $\mathbf{A}$ is given. The first subplot: the two curves of the arithmetic means $\bar{\alpha}^{(j)}$ ("o") and geometric means $\tilde{\alpha}^{(j)}$ ("+"). The second subplot: two curves of $P(\mathbf{s}^{(0)} = \mathbf{s}^{(1)}; \alpha_1^{(j)}, \ldots, \alpha_9^{(j)}, \mathbf{A})$ (the dotted line with "o") and $\mathrm{pro}(j; \alpha_1^{(j)}, \ldots, \alpha_9^{(j)}, \mathbf{A})$ (the solid line with "+"). The third subplot: three curves of $P(\mathbf{s}^{(0)} = \mathbf{s}^{(1)}; \alpha_1^{(j)}, \ldots, \alpha_9^{(j)}, \mathbf{A})$ (the dotted line with "o"), $\bar{\mathrm{pro}}(j; \bar{\alpha}^{(j)}, \mathbf{A})$ (the dashed–dotted line with "+") and $\bar{\mathrm{pro}}(j; \tilde{\alpha}^{(j)}, \mathbf{A})$ (the solid line with "+"). The second row is for the case, when $\mathbf{A}$ is random. The first subplot: two curves of $\bar{\alpha}^{(j)}$ ("o") and $\tilde{\alpha}^{(j)}$ ("+"). The second subplot: three curves of $\tilde{\mathrm{pro}}(j; \alpha_1^{(s)}, \ldots, \alpha_m^{(s)})$ (the dotted line with "o"), $\tilde{\mathrm{pro}}(j; \bar{\alpha}^{(j)})$ (the dashed–dotted line with "+") and $\tilde{\mathrm{pro}}(j; \tilde{\alpha}^{(j)})$ (the solid line with "+"). The third subplot: three curves of $\tilde{\mathrm{pro}}(j; \alpha_1^{(s)}, \ldots, \alpha_m^{(s)})$ (the dotted line with "o"), $\mathrm{pro}_m(j; \bar{\alpha}^{(j)})$ (the dashed–dotted line with "+") and $\mathrm{pro}_m(j; \tilde{\alpha}^{(j)})$ (the solid line with "+").

and $\tilde{\alpha}^{(j)}$ (denoted by "+"). The curves of the probabilities $P(\mathbf{s}^{(0)} = \mathbf{s}^{(1)}; \alpha_1^{(j)}, \ldots, \alpha_9^{(j)}, \mathbf{A})$ (dotted line with "o") and $\mathrm{pro}(j; \alpha_1^{(j)}, \ldots, \alpha_9^{(j)}, \mathbf{A})$ (solid line with "+") are shown in the second subplot. From this subplot, we can see that the two curves fit well. This demonstrates the validity of the probability estimation of (2).

The third subplot shows three curves of the probabilities $P(\mathbf{s}^{(0)} = \mathbf{s}^{(1)}; \alpha_1^{(j)}, \ldots, \alpha_9^{(j)}, \mathbf{A})$ (dotted line with "o"), $\bar{\mathrm{pro}}(j; \bar{\alpha}^{(j)}, \mathbf{A})$ (dashed–dotted line with "+"), and $\bar{\mathrm{pro}}(j; \tilde{\alpha}^{(j)}, \mathbf{A})$ (solid line with "+"). From this subplot, we can see the following: i) $P(\mathbf{s}^{(0)} = \mathbf{s}^{(1)}; \alpha_1^{(j)}, \ldots, \alpha_9^{(j)}, \mathbf{A})$ is close to $\bar{\mathrm{pro}}(j; \bar{\alpha}^{(j)}, \mathbf{A})$; and ii) the probability $\bar{\mathrm{pro}}(j; \tilde{\alpha}^{(j)}, \mathbf{A})$ is smaller than both $P(\mathbf{s}^{(0)} = \mathbf{s}^{(1)}; \alpha_1^{(j)}, \ldots, \alpha_9^{(j)}, \mathbf{A})$ and $\bar{\mathrm{pro}}(j; \bar{\alpha}^{(j)}, \mathbf{A})$.

ii) Now we consider the case where the mixing matrix $\mathbf{A}$ is random.

First, we randomly generate ten sparsity vectors $\alpha^{(j)} = [\alpha_1^{(j)}, \ldots, \alpha_9^{(j)}]$, $j = 1, \ldots, 10$, using the uniform distribution in $[0, 1]$.

Second, we estimate the probability $P(\mathbf{s}^{(0)} = \mathbf{s}^{(1)}; \alpha_1^{(s)}, \ldots, \alpha_m^{(s)})$ using the sampling method. The estimates are denoted as $\tilde{\mathrm{pro}}(j; \alpha_1^{(s)}, \ldots, \alpha_m^{(s)})$.

Third, for each $\bar{\alpha}^{(j)}$, we randomly generate 3000 matrices $\mathbf{A}^{(k)} \in R^{7\times 9}$ and 30 000 source vectors $\mathbf{s}^{(k)} \in R^9$. Suppose that among the 3000 columns, there are $\tilde{L}_j$ vectors recovered. Then, we have an estimate of the true probability $\tilde{\mathrm{pro}}(j; \bar{\alpha}^{(j)}) = \tilde{L}_j/3000$. Using $\tilde{\alpha}^{(j)}$ to replace $\bar{\alpha}^{(j)}$ $(j = 1, \ldots, 10)$, we perform similar simulation and obtain another probability estimate $\tilde{\mathrm{pro}}(j; \tilde{\alpha}^{(j)})$.

Fourth, we randomly generate ten mixing matrices $\mathbf{A}^{(k)} \in R^{7\times 9}$ $(k = 1, \ldots, 10)$. For each pair of $\mathbf{A}^{(k)}$ and $\bar{\alpha}^{(j)}$, we estimate the probability $\bar{\mathrm{pro}}(j; \bar{\alpha}^{(j)}, \mathbf{A}^{(k)})$ by the sampling method as in i). We calculate the average probabilities

$$\mathrm{pro}_m\left(j; \bar{\alpha}^{(j)}\right) = \frac{1}{10}\sum_{k=1}^{10} \bar{\mathrm{pro}}\left(j; \bar{\alpha}^{(j)}, \mathbf{A}^{(k)}\right), \quad j = 1, \ldots, 10. \tag{11}$$

Using $\tilde{\alpha}^{(j)}$ to replace $\bar{\alpha}^{(j)}$, we perform similar estimation and obtain the average probabilities $\mathrm{pro}_m(j; \tilde{\alpha}^{(j)})$, $j = 1, \ldots, 10$.

The results are shown in the second row of Fig. 1. The first subplot shows $\bar{\alpha}^{(j)}$ ("o") and $\tilde{\alpha}^{(j)}$ ("+"). In the second subplot, the dotted line with "o" represents the probabilities $\tilde{\mathrm{pro}}(j; \alpha_1^{(s)}, \ldots, \alpha_m^{(s)})$. The dotted line with "+" represents the probabilities $\tilde{\mathrm{pro}}(j; \bar{\alpha}^{(j)})$. The solid line with "+" represents the probabilities $\tilde{\mathrm{pro}}(j; \tilde{\alpha}^{(j)})$. In the third subplot, the dotted line with "o" represents the probabilities $\tilde{\mathrm{pro}}(j; \alpha_1^{(s)}, \ldots, \alpha_m^{(s)})$. The dotted line with "+" represents the probabilities $\mathrm{pro}_m(j; \bar{\alpha}^{(j)})$. The solid line with "+" represents the probabilities $\mathrm{pro}_m(j; \tilde{\alpha}^{(j)})$. From the two subplots, we can see the following: i) the probability $P(\mathbf{s}^{(0)} = \mathbf{s}^{(1)}; \alpha_1^{(s)}, \ldots, \alpha_m^{(s)})$ can be approximated by $P(\mathbf{s}^{(0)} = \mathbf{s}^{(1)}; \bar{\alpha}^{(s)})$; ii) $P(\mathbf{s}^{(0)} = \mathbf{s}^{(1)}; \alpha_1^{(s)}, \ldots, \alpha_m^{(s)})$ also can be approximated by $(1/N)\sum_{k=1}^{N} P(\mathbf{s}^{(0)} = \mathbf{s}^{(1)}; \bar{\alpha}^{(s)}, \mathbf{A}^{(k)})$, where $\mathbf{A}^{(k)}$, $k = 1, \ldots, N$, are N samples of random $\mathbf{A}$; and iii) $P(\mathbf{s}^{(0)} = \mathbf{s}^{(1)}; \tilde{\alpha}^{(s)})$ is less than $P(\mathbf{s}^{(0)} = \mathbf{s}^{(1)}; \bar{\alpha}^{(s)})$.

III. Sparsity Degree of Mixtures

In this section, we first discuss the sparsity degree of the observed mixture signal, then establish the relationship between the equivalence probability and this sparsity degree, and illustrate application of this relationship.

First, for a given 0-norm solution with sparsity degree vector $[\alpha_1^{(s)}, \ldots, \alpha_m^{(s)}]$, we suppose that the events $\{s_j^{(0)} = 0\}$, $j = 1, \ldots, m$, are independent. Since the basis matrix and the 0-norm solution in the model (1) are generally arbitrary, the equality $x_i = 0$ for any fixed i implies that $s_j^{(0)} = 0$ for $j = 1, \ldots, m$ (with probability one). Then, we have

$$\begin{aligned} P(x_i = 0) &= P\left(s_j^{(0)} = 0, j = 1, \ldots, m\right) \\ &= \prod_{j=1}^{m} P\left(s_j^{(0)} = 0\right) \\ &= \prod_{j=1}^{m} \alpha_j^{(s)} \end{aligned} \tag{12}$$

where $i = 1, \ldots, n$.

From (12), we can see that each x_i $(i = 1, \ldots, n)$ has the same sparsity degree denoted as $\alpha^{(x)}$, i.e.,

$$\alpha^{(x)} = \prod_{j=1}^{m} \alpha_j^{(s)}. \tag{13}$$

Remark 2: In model (1), suppose that we have a sufficient number of signal vectors $\mathbf{x}$ (we use them as column vectors to construct a matrix $\mathbf{X}$). Then, we have two methods to estimate $\alpha^{(x)}$. The first method is to estimate $\alpha^{(x)}$ using the norms of all columns of $\mathbf{X}$; the second method is to estimate $\alpha^{(x)}$ using a row

TABLE I
SPARSITY DEGREES $\alpha^{(s)}(n, m, p_0)$ FOR 0-NORM SOLUTIONS ($p_0 = 0.95$)

$n \setminus m$	7	8	9	10	11	12	13	14	15
6	0.76	0.82	0.88	0.90	0.92	0.93	0.93	0.95	0.96
7		0.71	0.82	0.85	0.88	0.90	0.92	0.93	0.94
8			0.70	0.77	0.83	0.86	0.89	0.91	0.92
9				0.66	0.75	0.80	0.85	0.88	0.90
10					0.64	0.73	0.80	0.83	0.85
11						0.63	0.72	0.78	0.82
12							0.61	0.71	0.76
13								0.58	0.68
14									0.56

TABLE II
SPARSITY DEGREES $\alpha^{(x)}(n, m, p_0)$ FOR SIGNAL VECTORS $\mathbf{x}$ ($p_0 = 0.95$)

$n \setminus m$	7	8	9	10	11	12	13	14	15
6	0.15	0.20	0.32	0.35	0.40	0.42	0.39	0.49	0.54
7		0.07	0.17	0.20	0.25	0.28	0.34	0.36	0.40
8			0.04	0.07	0.13	0.16	0.22	0.27	0.29
9				0.02	0.04	0.07	0.12	0.17	0.21
10					0.01	0.02	0.06	0.07	0.09
11						0.004	0.01	0.03	0.05
12							0.002	0.01	0.02
13								0.0005	0.003
14									0.0002

vector in $\mathbf{X}$. For model (1), the two methods give little difference in the result. If model (1) is not strictly true (e.g., when there is noise), we suggest the second one. This is because in such a case, a signal vector $\mathbf{x}$ is seldom equal to a zero vector even if $s_j = 0$ for $j = 1, \ldots, m$.

From (13), if $\alpha_1^{(s)} = \cdots = \alpha_m^{(s)} = \alpha^{(s)}$, then $\alpha^{(s)} = (\alpha^{(x)})^{1/m}$; otherwise, the geometric mean $\tilde{\alpha}^{(s)} = (\alpha^{(x)})^{1/m}$. In the following, we only consider the case in which all entries of the 0-norm solution have the same sparsity degree $\alpha^{(s)}$.

From (7), (8), and (13), we have the following corollary.

Corollary 2: For model (1), suppose that the sparsity degree of signal $\mathbf{x}$ is $\alpha^{(x)}$, and that all entries of the 0-norm solution have the same sparsity degree $\alpha^{(s)}$.

1) For a given basis matrix $\mathbf{A}$, we have

$$P\left(\mathbf{s}^{(0)} = \mathbf{s}^{(1)}; \alpha^{(x)}, \mathbf{A}\right) = \sum_{j=0}^{m} C_m^j \left(1 - \left(\alpha^{(x)}\right)^{\frac{1}{m}}\right)^j \times \left(\alpha^{(x)}\right)^{\left(1-\frac{j}{m}\right)} \frac{q_j}{2^j C_m^j} \quad (14)$$

where q_j is the number of sign vectors in T_j, which can be recovered by solving (P_1').

2) For a random basis matrix, we have

$$P\left(\mathbf{s}^{(0)} = \mathbf{s}^{(1)}; \alpha^{(x)}\right) = \sum_{j=0}^{m} C_m^j \left(1 - \left(\alpha^{(x)}\right)^{\frac{1}{m}}\right)^j \times \left(\alpha^{(x)}\right)^{\left(1-\frac{j}{m}\right)} P(n, m, j). \quad (15)$$

Equations (14) and (15) can be used to analyze the performance of BSS. Suppose that the mixing matrix has been estimated. We first estimate the sparsity degree of the available mixtures, then calculate the recoverability probability by (14), which is our estimate of the true recoverability probability. In fact, the true recoverability probability should be larger than this estimate since the sparsity degree of the mixtures corresponds to the geometric mean of the sparsity degrees of the sources.

In real-world applications, we set a threshold p_0 (e.g., 0.95) for the equivalence probability. Using (7) and (8) for a given basis matrix and a random basis matrix, respectively, we search an $\alpha^{(s)}$ in $[0, 1]$ such that the equivalence probability $P(\mathbf{s}^{(0)} = \mathbf{s}^{(1)}; \alpha^{(s)}, \mathbf{A})$ [or $P(\mathbf{s}^{(0)} = \mathbf{s}^{(1)}; \alpha^{(s)})$] just exceeds p_0. In this way, we determine the corresponding sparsity degree constraint of the 0-norm solution. In this paper, we consider the case in which $\mathbf{A}$ is random and perform a search of $\alpha^{(s)}$ in $\{0.01, 0.02, \ldots, 1\}$. Hereafter, this sparsity degree constraint is denoted as $\alpha^{(s)}(n, m, p_0)$, which is related to the dimension number n of signal vectors, the number m of bases, and the equivalence probability threshold p_0.

Under the equivalence probability threshold of 0.95, Table I shows the corresponding sparsity degrees $\alpha^{(s)}(n, m, p_0)$ for $n = 6, \ldots, 14$, and $m = n + 1, \ldots, 15$.

Using the sparsity degree constraints in Table I and (13), we can further estimate the corresponding sparsity degrees for signal vector $\mathbf{x}$, which are denoted as $\alpha^{(x)}(n, m, p_0)$. Table II shows the sparsity degrees $\alpha^{(x)}(n, m, p_0)$ corresponding to the equivalence probability threshold of 0.95.

Tables I and II provide a guide on how to guarantee the recoverability probability of the 0-norm solution. There are two cases. 1) All entries of the 0-norm solution have the same sparsity degree $\alpha^{(s)}$. If a threshold of recoverability probability (0.95) is imposed, then a corresponding constraint of the sparsity degree of the 0-norm solution in Table I (or a corresponding constraint of the sparsity degree of the mixtures in Table II) should be satisfied. In real-world applications, this constraint may not be satisfied in the time domain. However, it can be satisfied through a time-frequency transformation (e.g., wavelet packet transformation) in many cases [19], [22], [23]. 2) The entries of the 0-norm solution have different sparsity degrees. If the average sparsity degree (i.e., the arithmetic mean of the sparsity degree vector) of the sources is estimated, Table I can be used to analyze the performance of BSS based on (9). Otherwise, a sparsity degree of mixtures in Table II corresponds to the geometric mean of a sparsity degree vector of the 0-norm solution. If the sparsity degree constraint of mixtures in Table II is satisfied, then from (10) the recoverability probability is larger than 0.95. Thus, Tables I and II are still useful in this case.

In the next section, we will illustrate how to use the above sparsity degree constraints in Tables I and II to guarantee a satisfactory performance of BSS.

IV. Noisy Case

In this section, we will show that the results obtained in the previous sections (i.e., the estimates of the equivalence probability and the two relationships) still hold when the noise is sufficiently small. For simplicity, we only consider the case in which all entries of the 0-norm solution have the same sparsity degree. The case in which the entries of the 0-norm solution have different sparsity degrees can be discussed similarly.

For the noisy case, the optimization problem (P_1') becomes

$$(P_1'') \quad \min \sum_{i=1}^{m} |s_i|, \text{ s.t. } \mathbf{A}\mathbf{s} = \mathbf{x}_v^*$$

where $\mathbf{x}_v^* = \mathbf{A}\mathbf{s}^{(0)} + \mathbf{v}$, $\mathbf{s}^{(0)} \in \mathbf{R}^m$ being the 0-norm solution in the noiseless case and $\mathbf{v} \in \mathbf{R}^n$ a noise vector. The solution of (P_1'') is denoted as $\mathbf{s}_v^{(1)}$.

For a small positive constant θ, if

$$\left\|\mathbf{s}_v^{(1)} - \mathbf{s}^{(0)}\right\|_1 < \theta \tag{16}$$

then we say that $\mathbf{s}^{(0)}$ can be recovered by solving (P_1'') in the noisy case.

We first consider the case where the mixing matrix $\mathbf{A}$ is given or estimated.

According to the discussion in [11], the solution of (P_1') is robust to noise. Precisely speaking, for a given $\mathbf{A}$, there exists a $M > 0$, such that

$$\left\|\mathbf{s}_v^{(1)} - \mathbf{s}^{(1)}\right\|_1 < M\|\mathbf{v}\|_1 \tag{17}$$

where $\mathbf{s}_v^{(1)}$ and $\mathbf{s}^{(1)}$ are the solutions of (P_1'') and (P_1'), respectively.

Suppose that the noise is sufficiently small, such that $M\|\mathbf{v}\|_1 < \theta$. From (17), we have

$$\left\|\mathbf{s}_v^{(1)} - \mathbf{s}^{(1)}\right\|_1 < \theta. \tag{18}$$

Define a conditional probability

$$P\left(\left\|\mathbf{s}^{(1)} - \mathbf{s}^{(0)}\right\|_1 \leq 2\theta; \mathbf{A}, \mathbf{s}^{(1)} \neq \mathbf{s}^{(0)}\right) = \beta_0 \tag{19}$$

where β_0 is a positive constant.

The above probability β_0 is small due to the following two reasons: 1) θ is a small positive constant and 2) when the 1-norm solution $\mathbf{s}^{(1)}$ is not equal to the 0-norm solution $\mathbf{s}^{(0)}$, the difference $|\mathbf{s}^{(1)} - \mathbf{s}^{(0)}|$ is generally larger than the small positive constant 2θ. This is because there are infinite number of solutions of the constraint equations of (P_1') including $\mathbf{s}^{(0)}$ and $\mathbf{s}^{(1)}$, and $\mathbf{s}^{(1)}$ has the smallest 1-norm among these solutions.

From (18) and (19), it can be proved that (see Appendix II)

$$P\left(\left\|\mathbf{s}_v^{(1)} - \mathbf{s}^{(0)}\right\|_1 < \theta; \mathbf{A}\right) \approx P\left(\mathbf{s}^{(1)} = \mathbf{s}^{(0)}; \mathbf{A}\right) \tag{20}$$

where $P(\|\mathbf{s}_v^{(1)} - \mathbf{s}^{(0)}\|_1 < \theta; \mathbf{A})$ is the recoverability probability of $\mathbf{s}^{(0)}$ in the noisy case and $P(\mathbf{s}^{(1)} = \mathbf{s}^{(0)}; \mathbf{A})$ is the recoverability probability of $\mathbf{s}^{(0)}$ in the noiseless case.

From the discussion in [16]

$$P\left(\mathbf{s}^{(0)} = \mathbf{s}^{(1)}; \left\|\mathbf{s}^{(0)}\right\|_0 = j, \mathbf{A}\right) = \frac{q_j}{2^j C_m^j} \tag{21}$$

where q_j is the number of sign vectors in T_j, which can be recovered by solving (P_1').

From (21) and (20), we have

$$P\left(\left\|\mathbf{s}_v^{(1)} - \mathbf{s}^{(0)}\right\|_1 < \theta; \left\|\mathbf{s}^{(0)}\right\|_0 = j, \mathbf{A}\right) \approx \frac{q_j}{2^j C_m^j}. \tag{22}$$

Furthermore, from (22), we have

$$\begin{aligned}
&P\left(\left\|\mathbf{s}_v^{(1)} - \mathbf{s}^{(0)}\right\|_1 < \theta; \alpha^{(s)}, \mathbf{A}\right) \\
&\quad = \sum_{j=0}^{m} P\left(\left\|\mathbf{s}_v^{(1)} - \mathbf{s}^{(0)}\right\|_1 < \theta, \left\|\mathbf{s}^{(0)}\right\|_0 = j; \alpha^{(s)}, \mathbf{A}\right) \\
&\quad = \sum_{j=0}^{m} P\left(\left\|\mathbf{s}_v^{(1)} - \mathbf{s}^{(0)}\right\|_1 < \theta; \left\|\mathbf{s}^{(0)}\right\|_0 = j, \alpha^{(s)}, \mathbf{A}\right) \\
&P\left(\left\|\mathbf{s}^{(0)}\right\|_0 = j; \alpha^{(s)}, \mathbf{A}\right) \\
&\quad \approx \sum_{j=0}^{m} C_m^j \left(1 - \alpha^{(s)}\right)^j \left(\alpha^{(s)}\right)^{(m-j)} \frac{q_j}{2^j C_m^j}.
\end{aligned} \tag{23}$$

From the above analysis, we conclude that, for a given or estimated mixing matrix $\mathbf{A}$, the estimate of equivalence probability in (21) (see Appendix I) and the relationship (7) between sparsity degree of the 0-norm solution and the equivalence probability are still effective when the noise is sufficiently small.

When $\mathbf{A} \in \mathbf{R}^{n \times m}$ is random, our simulations in Example 2 show that (8) still holds when the noise is small, i.e.,

$$\begin{aligned}
P\left(\left\|\mathbf{s}_v^{(1)} - \mathbf{s}^{(0)}\right\|_1 < \theta; \alpha^{(s)}\right) &\approx \sum_{j=0}^{m} C_m^j \left(1 - \alpha^{(s)}\right)^j \\
&\quad \times \left(\alpha^{(s)}\right)^{(m-j)} P(n, m, j)
\end{aligned} \tag{24}$$

where $P(n, m, j)$ $(j = 0, \ldots, m)$ is the probability that a 0-norm solution with j nonzeros can be recovered by solving (P_1') in noiseless case. The estimation of $P(n, m, j)$ was discussed in [17].

Unfortunately, until now, we cannot prove (24). This is mainly because (17) does not hold for all arbitrary mixing matrices.

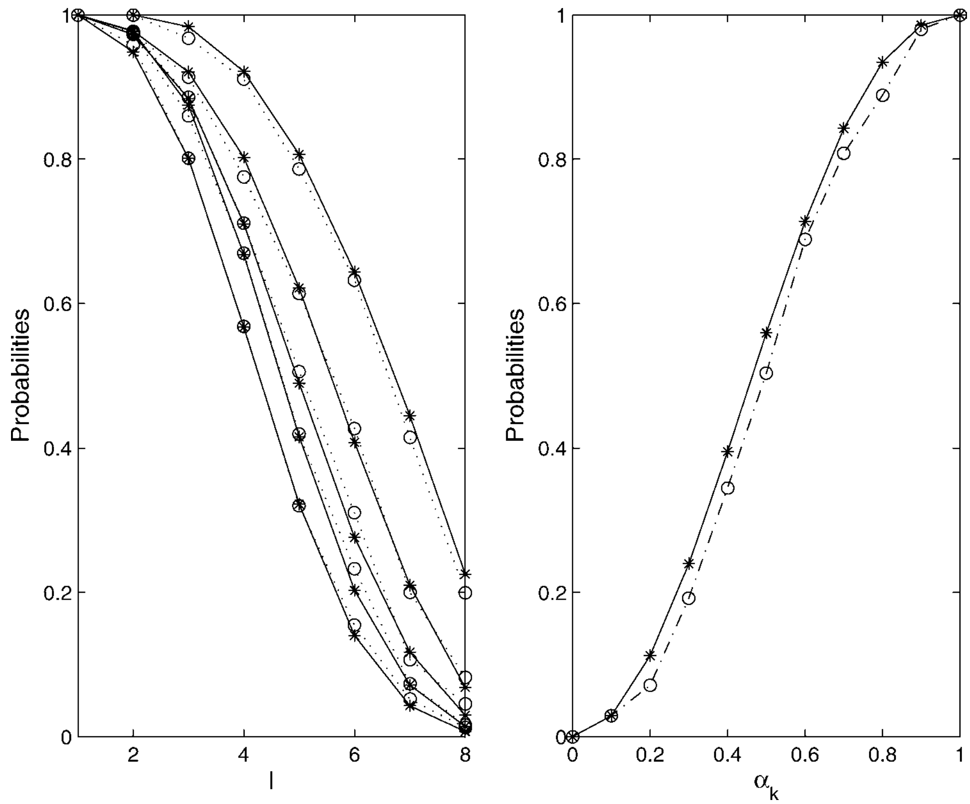

Fig. 2. Probabilities curves obtained in Example 2. In the left subplot, solid curves with "$*$": probability curves of $P(\mathbf{s}^{(0)} = \mathbf{s}^{(1)}; \|\mathbf{s}^{(0)}\|_0 = l, \mathbf{A}_i)$; dotted curves with "$o$": probability curves obtained in the noisy case. The five pairs of solid and dotted curves from top to bottom correspond to $m = 9, 10, 11, 12, 13$, respectively. x-axis label l represents the number of nonzeros of the source vectors. In the right subplot, the solid curve with "$*$" depicts our estimated probabilities $P(\mathbf{s}^{(0)} = \mathbf{s}^{(1)}; \alpha_k)$, $k = 1, \ldots, 11$, while the dashed curve with "o" represents the true equivalence probabilities obtained in the noisy case.

Furthermore, from (23) and (24), it is not difficult to extend (14) and (15) for the noisy case.

Example 2: This example contains two parts, in which we demonstrate the validity of (22) and (24), respectively.

i) First, we always set $n = 8$. For $m = 9, 10, 11, 12, 13$, we randomly generate five matrices denoted as $\mathbf{A}_i \in \mathbf{R}^{n \times m}$ $(i = 1, \ldots, 5)$ according to the uniform distribution in $[-1, 1]$.

For each m, the probabilities $P(\mathbf{s}^{(0)} = \mathbf{s}^{(1)}; \|\mathbf{s}^{(0)}\|_0 = l, \mathbf{A}_i)$, $l = 1, \ldots, 8$, are calculated based on (21) or by the sampling method.

Second, for each $\mathbf{A}_i$, we consider the noisy model (P_1'') with the zero mean Gaussian noise, where the average signal-to-noise ratio (SNR) is 18.8 dB. For each l, we randomly take 1500 source vectors (0-norm solutions) with l nonzeros, each of which is taken from the uniform distribution in $[-0.5, 0.5]$. By checking (16), we can find the number $u(m, l)$ of the source vectors which can be recovered by solving (P_1''). Note that θ in (16) is 0.1 here. In this way, we obtain the true equivalence probability estimate $u(m, l)/1500$ in the noisy case.

The left subplot of Fig. 2 shows our simulation results, from which (22) is demonstrated.

ii) We now demonstrate (24) by simulations. We set $n = 7$, $m = 9$, i.e., $\mathbf{A} \in \mathbf{R}^{7 \times 9}$. For $\alpha_k = (k - 1) * 0.1$, where $k = 1, \ldots, 11$, we calculate the probabilities $P(\mathbf{s}^{(0)} = \mathbf{s}^{(1)}; \alpha_k)$ according to (8). Next, for each k, we take 1500 pairs of mixing matrices and source vectors (0-norm solutions). All these mixing matrices are 7×9 dimensional and their entries are randomly valued in $[-1, 1]$. The 1500 source vectors are taken as follows:

$$s_q^* = 0, \ if \ \left|\tilde{s}_q^*\right| \le \frac{\alpha_j}{2}, \ s_q^* = \tilde{s}_q^*, \ if \ \left|\tilde{s}_q^*\right| > \frac{\alpha_j}{2} \qquad (25)$$

where $\tilde{\mathbf{s}}^* \in \mathbf{R}^9$ is drawn from the uniform distribution valued in $[-0.5, 0.5]$. From (25), we can see that the sparsity degree of the source vectors is α_j.

For each pair of source vectors and mixing matrices, we solve the linear programming problem (P_1'') with the zero mean Gaussian noise, where the average SNR is also 17.3 dB. By checking (16), with $\theta = 0.1$, we can determine whether a source vector is recovered. Suppose that m_k source vectors can be recovered in the noisy case, and thus, we obtain the ratio $\tilde{\text{pro}}(\alpha_k) = m_k/1500$, which reflects the true probability $P(\|\mathbf{s}_v^{(1)} - \mathbf{s}^0\|_1 < \theta)$ that the source vectors can be recovered in the noisy case under the distribution parameter α_k.

The curves of $P(\mathbf{s}^{(0)} = \mathbf{s}^{(1)}; \alpha_k)$ and $\tilde{\text{pro}}(\alpha_k)$ are shown in the right subplot of Fig. 2, which fit very well. Therefore, if the noise is sufficiently small, (24) still holds.

The parameter θ in (16) should be determined depending on real tasks. If θ is very small, then the SNR in the noisy model should be large, and vice versa. After θ is given, we can determine the approximate lower bound SNR by simulations as in Example 2. If the SNR is not sufficiently large, then the true probability estimated in the noisy case will be smaller than its corresponding theoretical estimate.

TABLE III
AVERAGE SPARSITY DEGREES OF SIGNALS IN THE TIME-FREQUENCY DOMAIN CORRESPONDING TO DIFFERENT NUMBERS OF THE LEVELS OF WPT

signals \ level	1	2	3	4	5	6	7	8	9	10
Speech sources	0.55	0.64	0.71	0.74	0.75	0.78	0.82	0.83	0.83	0.84
Speech mixtures	0.09	0.13	0.18	0.18	0.19	0.20	0.22	0.24	0.27	0.32

In this section, we have illustrated that the equivalence probability estimates (21) (presented in Appendix I) and the relationships (7) and (8) can be extended for the case when the noise is sufficiently small.

V. APPLICATION IN UNDERDETERMINED BSS

It is well known that sparse representation can be used in underdetermined BSS. The higher the sparsity degrees of the sources are, the better is the performance of BSS [16]. In this section, we illustrate the application of our results in underdetermined BSS.

We consider the following linear mixing model:

$$\mathbf{X} = \mathbf{A}\mathbf{S} \tag{26}$$

where $\mathbf{X} \in R^{n\times K}$ $(= [\mathbf{x}(1), \ldots, \mathbf{x}(K)])$ is an observed mixture matrix, $\mathbf{A} \in R^{n\times m}$ is an unknown mixing matrix, and $\mathbf{S} \in R^{m\times K}$ $(= [\mathbf{s}(1), \ldots, \mathbf{s}(K)])$ is an unknown source matrix. Each row of $\mathbf{X}$ is a mixture signal, while each row of $\mathbf{S}$ is a source. The task of BSS is to estimate $\mathbf{S}$ using $\mathbf{X}$. In this paper, we assume that $n < m$. Thus, BSS here is underdetermined.

We use a two-stage clustering-then-l^1-optimization approach for the underdetermined BSS [19]. If the unknown sources are sufficiently sparse, the two-stage sparse representation approach can be directly applied to $\mathbf{X}$. Based on the known mixture matrix $\mathbf{X}$, the mixing matrix $\mathbf{A}$ is estimated in the first stage, and the source matrix $\mathbf{S}$ is estimated in the second stage by solving the optimization problem (P_1). However, real-world sources may not be sparse. In this case, we first apply a wavelet packet transform (WPT) to all mixtures for producing sparsity, and then perform BSS in the time-frequency domain. (In this paper, we always use Daubechies WPT). Since WPT is linear, we have

$$\tilde{\mathbf{X}} = \mathbf{A}\tilde{\mathbf{S}} \tag{27}$$

where the ith row of $\tilde{\mathbf{X}}$ is obtained by applying WPT to the ith row of $\mathbf{X}$ $(i = 1, \ldots, n)$, the jth row of $\tilde{\mathbf{S}}$ is the transformed source corresponding to the jth row of $\mathbf{S}$ $(j = 1, \ldots, m)$. Based on the principle of transform sparsity [24], each transformed mixture signal or each transformed source is sparse.

After we estimate $\tilde{\mathbf{S}}$ through BSS in the time-frequency domain, we reconstruct the source matrix $\mathbf{S}$ by applying the inverse WPT to each row of the estimated $\bar{\mathbf{S}}$ (which denotes the estimate of $\tilde{\mathbf{S}}$). Denote the jth rows of $\tilde{\mathbf{S}}$ and $\bar{\mathbf{S}}$ as $\tilde{\mathbf{s}}_j$ and $\bar{\mathbf{s}}_j$, respectively, $(j = 1, \ldots, m)$. Although there generally exists an error between $\tilde{\mathbf{s}}_j$ and $\bar{\mathbf{s}}_j$, each entry of $\tilde{\mathbf{s}}_j$ has a corresponding estimate in $\bar{\mathbf{s}}_j$. Thus, we can apply the inverse WPT to $\bar{\mathbf{s}}_j$ to reconstruct the source $\mathbf{s}_j$. In recent years, the field of compressed sensing has received a lot of attention [24]–[26]. If only part of entries of $\tilde{\mathbf{s}}_j$ are estimated and the locations of these estimated entries in $\tilde{\mathbf{s}}_j$ are unknown, then compressed sensing method can be used to reconstruct the source $\mathbf{s}_j$. If the error between $\tilde{\mathbf{s}}_j$ and $\bar{\mathbf{s}}_j$ is considered, an interesting research problem may arise here: Is compressed sensing method useful here in reconstructing the source $\mathbf{s}_j$ using the nonzeros of $\bar{\mathbf{s}}_j$ and better than the inverse WPT in performance? This problem will be studied in our future work.

The procedure for estimating the mixing matrix $\mathbf{A}$ has been discussed in [16]. When the mixing matrix is estimated correctly (or sufficiently accurately estimated), we need to discuss the recoverability problem, which can be rephrased as the question: How is it possible for the 1-norm solution to be equal to the true source vector? As stated before, this recoverability problem can be understood as the equivalence problem between the 0-norm and 1-norm solutions. In the following example, we will illustrate how to use the sparsity degree constraint (corresponding to an equivalence probability threshold) shown in Tables I and II to guarantee a satisfactory performance of BSS, and how to achieve these sparsity degree constraints.

Example 3: In this example, natural speech signals are analyzed. Since natural speech signals are not sufficiently sparse, we first apply a wavelet packet transform (WPT) to all mixtures. Thus, we consider the model (27) in the following. We need to choose a suitable level number of the WPT such that the average sparsity degree of these transformed mixture signals is larger than $\alpha^{(x)}(n, m, p_0)$. BSS is then performed in the time-frequency domain. With this approach, we ensure that the recoverability probability constraint p_0 is satisfied in the analyzed domain and guarantee a satisfactory performance of BSS.

To determine the number of levels of the WPT, we perform the ten-level Daubechies WPT to eight speech sources and six mixtures of the eight sources, and estimate the probabilities (i.e., the sparse degrees) with which the time-frequency coefficients are equal to zero. If the absolute value of a coefficient is less than $0.004M$, where M is the maximum of absolute values of all the coefficients, then this coefficient is taken to be zero. By calculating the ratio of zeros in all coefficients, we estimate the probability with which the coefficients are equal to zero. The second row of Table III lists the ten average sparsity degrees of the eight sources corresponding to ten levels of WPT, while the third row lists the tne average sparsity degrees of the six mixtures.

We now consider BSS of speech mixtures, in which the mixing matrix is 6×8 dimensional. From Table II, the sparsity degree of mixtures should be larger than 0.2 such that the recoverability probability is larger than 0.95. From Table III, we find that if we apply the seven-level WPT to the mixtures, then the average sparsity degree of the transformed mixtures is 0.22. That is, we can first perform the seven-level WPT to all mixtures to produce sparsification, and then carry out BSS in the time-frequency domain. Subsequently, the sources in the time domain are reconstructed by applying the inverse WPT to the transformed sources estimated in the time-frequency domain. In Fig. 3, the eight original speech sources are shown in the first row, the six mixtures are shown in the second row, and the recovered sources based on the seven-level WPT are

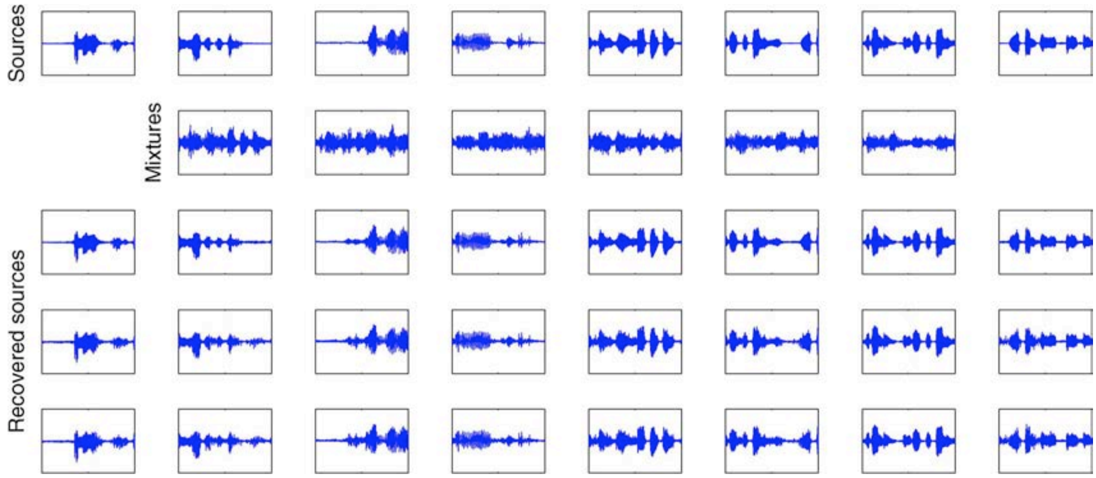

Fig. 3. Separation results in Example 3. The first row: eight speech sources; the second row: six mixtures; the third row: eight recovered sources based on a seven-level WPT; the fourth row: eight recovered sources based on a one-level WPT; the fifth row: eight recovered sources based on sparse representation in the time domain (without WPT).

shown in the third row. We compare the eight original sources and the eight recovered sources, respectively, and obtain the average SNR 15.9 dB, where the difference (or error) between the original sources and the recovered sources is taken as noise. Therefore, we find that the sparsity degree of mixtures produced by the seven-level Daubechies WPT is sufficient for blind separation of the speech sources in this example.

For the purpose of comparison, we also perform the one-level WPT to all the mixtures to produce sparsification, and then carry out BSS in the time-frequency domain. The fourth row of Fig. 3 shows our separation results (recovered sources). We compare the eight original sources and these eight recovered sources (shown in the fourth row), and obtain the average SNR 9.6 dB. Furthermore, without performing any WPT to the mixtures, we directly perform source separation in the time domain using sparse representation. The fifth row of Fig. 3 shows our separation results (recovered sources). We compare the eight original sources and these eight recovered sources (shown in the fifth row), and obtain the average SNR 8.7 dB. Therefore, the performance of separation based on the one-level Daubechies WPT is significantly worse than that based on the seven-level Daubechies WPT. This is because the sparsity degree of mixtures produced by the seven-level Daubechies WPT is not sufficient for blind separation of the speech sources, and so does the performance of the separation in the time domain.

In this example, although the sparsity degrees for the eight speech sources are different in general, we still use the sparsity degree constraint on the mixtures in Table II. As stated in Section III, this sparsity degree constraint on the mixtures corresponds to the geometric mean of the sparsity degrees of the sources. Furthermore, the recoverability probability for each source vector is larger than 0.95. Additionally, noise is not considered in Example 3. The readers can deal with the noisy model (P_1'') similarly as above.

VI. Conclusion

In the sparse representation of signals, we often focus on the two sparse solutions, the 0-norm and 1-norm solutions. The first solution is the most sparse, while the second solution is relatively easier to obtain. In this paper, the equivalence problem of these two sparse solutions was discussed within a probabilistic framework. We first defined the sparsity degree of a random variable or a signal sequence. The equivalence probability estimation is discussed when all the entries of the 0-norm solution have different sparsity degrees. For a given basis matrix as well as a random basis matrix, we established two relationships. The first is between the sparsity degree vector of the 0-norm solution and the equivalence probability, while the second is between the sparsity degree of the observed signal vector (mixture) and the equivalence probability. The computational burden for estimating the equivalence probability increases exponentially as m (the number of the dimension of the 0-norm solution) increases. The sampling method can be used to resolve this computational complexity problem. We further show that these theoretical results are still effective when the noise is sufficiently small. Finally, we illustrate how to use the two relationships to guarantee a satisfactory performance in BSS.

Appendix I
The Sampling Method for Estimating the Equivalence Probability

In this appendix, we mainly discuss the empirical estimation of two equivalence probabilities presented in [16] and [17], respectively, when m is large.

When all entries of $\mathbf{A} \in \mathbf{R}^{n\times m}$ are drawn from a distribution (e.g., a uniform distribution valued in $[-1, 1]$), then the equivalence probability depends on n, m, and the number l of nonzeros of $\mathbf{s}^{(0)}$. Hence, we denote

$$P(n,m,l) = P\left(\mathbf{s}^{(1)} = \mathbf{s}^{(0)}; \left\|\mathbf{s}^{(0)}\right\|_0 = l\right) \tag{28}$$

where $P(\mathbf{s}^{(1)} = \mathbf{s}^{(0)}; \|\mathbf{s}^{(0)}\|_0 = l)$ is a conditional probability (with the condition $\|\mathbf{s}^{(0)}\|_0 = l$).

Let $T_l = \{\mathbf{t} = [t_1, \ldots, t_m]^T | t_j \in \{-1, 0, 1\}, \sum_{j=1}^m |t_j| = l\}$, which is a set of sign vectors with l nonzero entries. Obviously, there are $2^l C_m^l$ vectors in T_l, where $C_m^l = m!/(m-l)!l!$.

Lemma 1 [16]: For a given basis matrix A, suppose that there are q_l sign column vectors in T_l that can be recovered by solving (P_1'). Then, we have

$$P\left(\mathbf{s}^{(0)} = \mathbf{s}^{(1)}; \left\|\mathbf{s}^{(0)}\right\|_0 = l, \mathbf{A}\right) = \frac{q_l}{2^l C_m^l} \tag{29}$$

where $l = 1, 2, \ldots, n$.

When the basis matrix is given, we can check how many sign vectors with l nonzeros can be recovered by solving (P_1') and then calculate the equivalence probability using (29). To calculate the probability in (29), we need to check $2^l C_m^l$ sign vectors and find the number q_l of sign vectors which can be recovered by solving (P_1'). Thus, when m increases, the computational burden for estimating the probability in (29) will increase exponentially. The following theorem provides a solution to this problem.

Theorem 2: Suppose that M_l sign vectors are randomly taken from T_l, where M_l is a large positive integer ($M_l \ll 2^l * C_m^l$ when m is large), and K_l of the M_l sign vectors can be recovered by solving linear problem (P_1'). Then

$$P\left(\mathbf{s}^{(0)} = \mathbf{s}^{(1)}; \left\|\mathbf{s}^{(0)}\right\|_0 = l, \mathbf{A}\right) = \frac{q_l}{2^l C_m^l} \simeq \frac{K_l}{M_l}. \tag{30}$$

Proof: Let $T_l = T_{l1} \cup T_{l2}$, where T_{l1} contains the q_l vectors that can be recovered by solving (P_1'), $T_{l2} = T_l \setminus T_{l1}$. For a sign vector $\mathbf{t}$ with l nonzero entries, we have

$$P(\mathbf{t} \in T_{l1}) = \frac{q_l}{2^l C_m^l}. \tag{31}$$

Now we define a sequence of random variables v_k using the set of sign vectors T_l

$$v_k = \begin{cases} 1, & \text{if } \mathbf{t}_k \in T_{l1} \\ 0, & \text{if } \mathbf{t}_k \in T_{l2} \end{cases} \tag{32}$$

where $k = 1, 2, \ldots, \mathbf{t}_k$ is a sign vector randomly taken from T_l.

From (31), we have $P(v_k = 1) = q_l/2^l C_m^l$, $P(v_k = 0) = 1 - (q_l/2^l C_m^l)$. Therefore, v_k, $k = 1, 2, \ldots$, are independent identically distributed random variables with the expected value $E(v_k) = q_l/2^l C_m^l$.

According to the law of large numbers (Bernoulli) in probability theory, the sample average $(1/M_l)\sum_{i=1}^{M_l} V_i$ converges towards the expected value $E(v_k)$, where V_i is a sample of the random variable v_i. Furthermore, the condition in this theorem implies that $(1/M_l)\sum_{i=1}^{M_l} V_i = K_l/M_l$. Thus, when M_l is sufficiently large

$$E(v_k) = \frac{q_l}{2^l C_m^l} \simeq \frac{K_l}{M_l}. \tag{33}$$

The theorem is proven.

From Theorem 2, to estimate the probability $P(\mathbf{s}^{(0)} = \mathbf{s}^{(1)}; \|\mathbf{s}^{(0)}\|_0 = l, \mathbf{A})$ when m is large, we need not check all $2^l C_m^l$ sign vectors. It is sufficient to randomly take M_l sign vectors from T_l, and determine how many of the M_l sign vectors which can be recovered by solving (P_1').

When the basis is random, we have the following approximation of the equivalence probability [17]:

$$P(n, m, l) \approx \frac{1}{k_0}\sum_{q=1}^{k_0} P\left(\mathbf{s}^{(0)} = \mathbf{s}^{(1)}; \left\|\mathbf{s}^{(0)}\right\|_0 = l, \mathbf{A}_q\right) \tag{34}$$

where $\mathbf{A}_1, \ldots, \mathbf{A}_{k_0}$ are k_0 random samples of $\mathbf{A}$, and k_0 is a small positive integer (5 in the simulation examples in [17]).

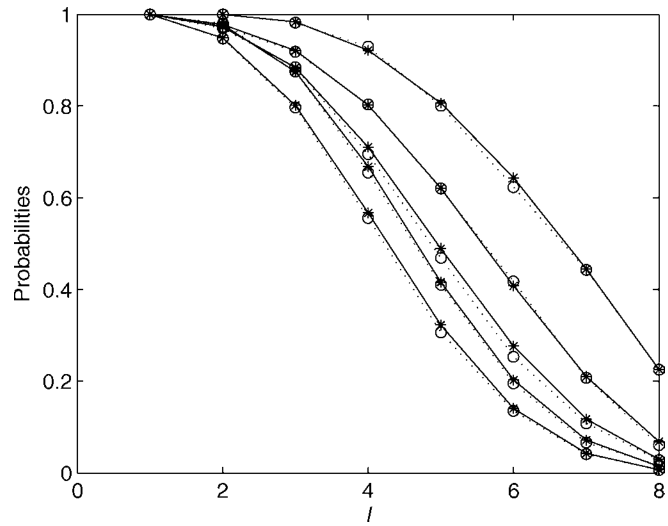

Fig. 4. Probabilities curves obtained in Example 1. Solid curves with "$*$": probability curves of $P(\mathbf{s}^{(0)} = \mathbf{s}^{(1)}; \|\mathbf{s}^{(0)}\|_0 = l, \mathbf{A}_i)$ obtained by (29); dotted curves with "o": probability curves obtained by a sampling method. The five pairs of solid and dotted curves from the top to the bottom correspond to $m = 9, 10, 11, 12, 13$, respectively. Axis label l represents the number of nonzeros of source vectors.

When m is large, as stated previously, the probability $P(\mathbf{s}^{(0)} = \mathbf{s}^{(1)}; \|\mathbf{s}^{(0)}\|_0 = l, \mathbf{A}_q)$ on the right-hand side of (34) can be estimated by (30). Therefore, when the basis is random, the equivalence probability $P(n, m, l)$ can still be estimated by the sampling method.

A problem may arise here. When m increases, does the number M_l of samples need to increase exponentially (just like the number of sign vectors)? From the proof of Theorem 2, M_l, which controls the precision in the approximation of (33), i.e., the precision of the sampling estimation, is related to the two-point distribution of v_k other than the size of T_l. Thus, there is no need for M_l increasing exponentially as m increases. In the following simulation Example 4, we will demonstrate this conclusion as well as the conclusion in Theorem 2.

Example 4: In this example, we always set $n = 8$. For $m = 9, 10, 11, 12, 13$, we randomly generate five matrices denoted as $\mathbf{A}_i \in \mathbf{R}^{n \times m}$ $(i = 1, \ldots, 5)$ according to a uniform distribution in $[-1, 1]$.

For each m, we first calculate the probabilities $P(\mathbf{s}^{(0)} = \mathbf{s}^{(1)}; \|\mathbf{s}^{(0)}\|_0 = l, \mathbf{A}_i)$ according to (29), $l = 1, \ldots, 8$. Note that $P(\mathbf{s}^{(0)} = \mathbf{s}^{(1)}; \|\mathbf{s}^{(0)}\|_0 = l, \mathbf{A}_i) = 0$ for $l > 8$. The number of sign vectors increases exponentially as m increases. This leads to an exponential increase of the computational burden for estimating these probabilities. For example, for $m = 9, 10$, $11, 12, 13$, we need to solve 19 170, 52 904, 135 674, 322 544, 714 194 optimization problems, respectively. This makes the estimates difficult to obtain for large m (e.g., $m > 15$).

From Theorem 2, for a given matrix $\mathbf{A}_i$, we can estimate these probabilities by a sampling method. For each l, we randomly take 3000 sign vectors, and find the number $k(m, l)$ of sign vectors which can be recovered by solving (P_1'). Then, we obtain the equivalence probability estimate $k(m, l)/3000$.

Fig. 4 shows our simulation results. From Fig. 4, we have two conclusions: 1) the probabilities estimated by the sampling method reflect their corresponding theoretical values well; 2) the number M_l of samples in Theorem 2 need not increase exponentially as m increases. Thus, the computational complexity in (29) can be avoided.

Appendix II
The Proof of (20)

We now prove (20) using four steps.

First, (19) and its condition imply that

$$P\left(\left\|\mathbf{s}^{(1)}-\mathbf{s}^{(0)}\right\|_1 > 2\theta; \mathbf{A}, \mathbf{s}^{(1)} \neq \mathbf{s}^{(0)}\right) > 1-\beta_0. \tag{35}$$

That is

$$P\left(\left\|\mathbf{s}^{(1)}-\mathbf{s}^{(0)}\right\|_1 \le 2\theta; \mathbf{A}, \mathbf{s}^{(1)} \neq \mathbf{s}^{(0)}\right) < \beta_0. \tag{36}$$

Next

$$\begin{aligned} &P\left(\left\|\mathbf{s}^{(1)}-\mathbf{s}^{(0)}\right\|_1 \le 2\theta; \mathbf{A}\right) \\ &\quad= P\left(\left\|\mathbf{s}^{(1)}-\mathbf{s}^{(0)}\right\|_1 \le 2\theta, \mathbf{s}^{(1)} = \mathbf{s}^{(0)}; \mathbf{A}\right) \\ &\qquad+ P\left(\left\|\mathbf{s}^{(1)}-\mathbf{s}^{(0)}\right\|_1 \le 2\theta, \mathbf{s}^{(1)} \neq \mathbf{s}^{(0)}; \mathbf{A}\right) \\ &\quad= P\left(\mathbf{s}^{(1)} = \mathbf{s}^{(0)}; \mathbf{A}\right) \\ &\qquad+ P\left(\left\|\mathbf{s}^{(1)}-\mathbf{s}^{(0)}\right\|_1 \le 2\theta, \mathbf{s}^{(1)} \neq \mathbf{s}^{(0)}; \mathbf{A}\right) \\ &\quad= P\left(\mathbf{s}^{(1)} = \mathbf{s}^{(0)}; \mathbf{A}\right) \\ &\qquad+ P\left(\left\|\mathbf{s}^{(1)}-\mathbf{s}^{(0)}\right\|_1 \le 2\theta; \mathbf{A}, \mathbf{s}^{(1)} \neq \mathbf{s}^{(0)}\right) \\ &\qquad\times P\left(\mathbf{s}^{(1)} \neq \mathbf{s}^{(0)}; \mathbf{A}\right) \\ &\quad\le P\left(\mathbf{s}^{(1)} = \mathbf{s}^{(0)}; \mathbf{A}\right) + \beta_0 \end{aligned} \tag{37}$$

where the last inequality is based on (36) and that $P(\mathbf{s}^{(1)} \neq \mathbf{s}^{(0)}; \mathbf{A}) \le 1$.

Third, obviously

$$\begin{aligned} \left\|\mathbf{s}_v^{(1)}-\mathbf{s}^{(0)}\right\|_1 &= \left\|\mathbf{s}_v^{(1)}-\mathbf{s}^{(1)}+\mathbf{s}^{(1)}-\mathbf{s}^{(0)}\right\|_1 \\ &\ge \left\|\mathbf{s}^{(1)}-\mathbf{s}^{(0)}\right\|_1 - \left\|\mathbf{s}_v^{(1)}-\mathbf{s}^{(1)}\right\|_1. \end{aligned} \tag{38}$$

From (18) and (38), if $\|\mathbf{s}_v^{(1)}-\mathbf{s}^{(0)}\|_1 < \theta$, then $\|\mathbf{s}^{(1)}-\mathbf{s}^{(0)}\|_1 \le 2\theta$. Furthermore, considering (37), we have

$$\begin{aligned} P\left(\left\|\mathbf{s}_v^{(1)}-\mathbf{s}^{(0)}\right\|_1 \le \theta; \mathbf{A}\right) &\le P\left(\left\|\mathbf{s}^{(1)}-\mathbf{s}^{(0)}\right\|_1 \le 2\theta; \mathbf{A}\right) \\ &\le P\left(\mathbf{s}^{(1)} = \mathbf{s}^{(0)}; \mathbf{A}\right) + \beta_0. \end{aligned} \tag{39}$$

Fourth, if $\mathbf{s}^{(1)} = \mathbf{s}^{(0)}$, it follows from (18) that

$$\left\|\mathbf{s}_v^{(1)}-\mathbf{s}^{(0)}\right\|_1 = \left\|\mathbf{s}_v^{(1)}-\mathbf{s}^{(1)}\right\|_1 < \theta. \tag{40}$$

Thus

$$\begin{aligned} P\left(\mathbf{s}^{(1)} = \mathbf{s}^{(0)}; \mathbf{A}\right) &\le P\left(\left\|\mathbf{s}_v^{(1)}-\mathbf{s}^{(0)}\right\|_1 < \theta; \mathbf{A}\right) \\ &\le P\left(\mathbf{s}^{(1)} = \mathbf{s}^{(0)}; \mathbf{A}\right) + \beta_0 \end{aligned} \tag{41}$$

where the second inequality is based on (39).

Considering that β_0 in (41) is small, we have (20), i.e., the two recoverability probability $P(\mathbf{s}^{(1)} = \mathbf{s}^{(0)}; \mathbf{A})$ and $P(\|\mathbf{s}_v^{(1)}-\mathbf{s}^{(0)}\|_1 < \theta; \mathbf{A})$ are close to each other.

Acknowledgment

The authors would like to thank the anonymous reviewers for their constructive suggestions and they acknowledge their contributions to this paper.

References

[1] 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.

[2] B. A. Olshausen, P. Sallee, and M. S. Lewicki, "Learning sparse image codes using a wavelet pyramid architecture," in *Advances in Neural Information Processing Systems 13*. Cambridge, MA: MIT Press, 2001, pp. 887–893.

[3] M. S. Lewicki and T. J. Sejnowski, "Learning overcomplete representations," *Neural Comput.*, vol. 12, no. 2, pp. 337–365, 2000.

[4] K. K. Delgado, J. F. Murray, B. D. Rao, K. Engan, T. W. Lee, and T. J. Sejnowski, "Dictionary learning algorithms for sparse representation," *Neural Comput.*, vol. 15, pp. 349–396, 2003.

[5] R. Gribonval and M. Nielsen, "Sparse representations in unions of bases," *IEEE Trans. Inf. Theory*, vol. 49, no. 12, pp. 3320–3325, Dec. 2003.

[6] J. A. Tropp, "Algorithms for simultaneous sparse approximation, Part II: Convex relaxation," *EURASIP J. Appl. Signal Process.*, Special Issue on Sparse Approximations in Signal and Image Processing, 2005, accepted for publication.

[7] J. A. Tropp, A. C. Gilbert, S. Muthukrishnan, and M. J. Strauss, "Improved sparse approximation over quasi-incoherent dictionaries," in *Proc. IEEE Int. Conf. Image Process.*, Barcelona, Spain, Sep. 2003, pp. 37–40.

[8] J. A. Tropp, "Greed is good: Algorithmic results for sparse approximation," *IEEE Trans. Inf. Theory*, vol. 50, no. 10, Oct. 2004.

[9] R. Gribonval, R. Figueras, and P. Vandergheynst, "A simple test to check the optimality of sparse signal approximations," in *Proc. IEEE Conf. Acoust. Speech Signal Process.*, 2005, vol. V, pp. 717–720.

[10] D. L. Donoho and M. Elad, "Maximal sparsity representation via l^1 minimization," *Proc. Nat. Acad. Sci.*, vol. 100, pp. 2197–2202, 2003.

[11] Y. Q. Li, A. Cichocki, and S. Amari, "Analysis of Sparse representation and blind source separation," *Neural Comput.*, vol. 16, pp. 1193–1234, 2004.

[12] J.-J. Fuchs, "On sparse representations in arbitrary redundant bases," *IEEE Trans. Inf. Theory*, vol. 50, no. 6, pp. 1341–1344, Jun. 2004.

[13] D. L. Donoho, M. Elad, and V. N. Temlyakov, "Stable recovery of sparse overcomplete representations in the presence of noise," Feb. 2004, working draft.

[14] D. L. Donoho and X. Huo, "Uncertainty principles and ideal atomic decomposition," *IEEE Trans. Inf. Theory*, vol. 47, no. 7, pp. 2845–2862, Jul. 2001.

[15] D. L. Donoho, "For most large underdetermined systems of linear equations, the minimal l1-norm solution is also the sparsest solution," (Report), 2004 [Online]. Available: http://www-stat.stanford.edu/donoho/reports.html

[16] Y. Q. Li, A. Cichocki, S. Amari, 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.

[17] Y. Q. Li, S. Amari, A. Cichocki, and C. T. Guan, "Probability estimation for recoverability analysis of blind source separation based on sparse representation," *IEEE Trans. Inf. Theory*, vol. 52, no. 7, pp. 3139–3152, Jul. 2006.

[18] P. Bofill and M. Zibulevsky, "Underdetermined blind source separation using sparse representations," *Signal Process.*, vol. 81, no. 11, pp. 2353–2362, 2001.

[19] M. Zibulevsky and B. A. Pearlmutter, "Blind source separation by sparse decomposition," *Neural Comput.*, vol. 13, no. 4, pp. 863–882, 2001.

[20] T. W. Lee, M. S. Lewicki, M. Girolami, and T. J. Sejnowski, "Blind source separation of more sources than mixtures using overcomplete representations," *IEEE Signal Process. Lett.*, vol. 6, no. 4, pp. 87–90, Apr. 1999.

[21] M. Girolami, "A variational method for learning sparse and overcomplete representations," *Neural Comput.*, vol. 13, no. 11, pp. 2517–1532, 2001.

[22] P. Srinivasan and L. H. Jamieson, "High quality audio compression using an adaptive wavelet packet decomposition and psychoacoustic modeling," *IEEE Trans. Signal Process.*, vol. 46, no. 4, pp. 1085–1093, Apr. 1998.

[23] L. Daudet, "Sparse and structured decompositions of audio signals in overcomplete spaces," in *Proc. 7th Int. Conf. Digital Audio Effects*, Naples, Italy, Oct. 5–8, 2004, pp. 22–26.
[24] D. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289–1306, Apr. 2006.
[25] E. J. Candes and T. Tao, "Decoding by linear programming," *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 4203–4215, Dec. 2005.
[26] E. J. Candes and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?," *IEEE Trans. Inf. Theory*, vol. 52, no. 12, pp. 5406–5425, Dec. 2006.

Yuanqing Li (M'06) was born in Hunan Province, China, in 1966. He received the B.S. degree in applied mathematics from Wuhan University, Wuhan, China, in 1988, the M.S. degree in applied mathematics from South China Normal University, Guangzhou, China, in 1994, and the Ph.D. degree in control theory and applications from South China University of Technology, Guangzhou, China, in 1997.

Since 1997, he has been with South China University of Technology, where he became a Full Professor in 2004. In 2002–2004, he was a Researcher at the Laboratory for Advanced Brain Signal Processing, RIKEN Brain Science Institute, Saitama, Japan. In 2004–2008, he was a Research Scientist at the Laboratory for Neural Signal Processing, Institute for Inforcomm Research, Singapore. He is the author or coauthor of more than 60 scientific papers in journals and conference proceedings. His research interests include, blind signal processing, sparse representation, machine learning, brain–computer interface, brain signal analysis.

Andrzej Cichocki (M'96–SM'07) was born in Poland. He received the M.Sc. (with honors), Ph.D., and Habilitate Doctorate (Dr.Sc.) degrees, from Warsaw University of Technology, Warsaw, Poland, in 1972, 1975, and 1982, respectively, all in electrical engineering.

Since 1972, he has been with the Institute of Theory of Electrical Engineering and Electrical Measurements, Warsaw University of Technology, where he became a full Professor in 1991. He is the coauthor of three books: *Adaptive Blind Signal and Image Processing* (New York: Wiley, 2003 with Prof. S.-. Amari), *MOS Switched-Capacitor and Continuous-Time Integrated Circuits and Systems* (New York: Springer-Verlag, 1989, with Prof. R. Unbehauen), and *Neural Networks for Optimization and Signal Processing* (New York: Wiley/Teubner-Verlag, 1993/1994, with Prof. R. Unbehauen) and coauthor of more than one 150 papers. Two of his books have been translated to Chinese and other languages. He a few years spent at University Erlangen-Nuernberg (Germany) as Alexander Humboldt Research Fellow and Guest Professor working in the area very large scale integration (VLSI) of electronic circuits, artificial neural networks, and optimization. He conducted and realized several successful research projects. In 1996–1999 he worked as a Team Leader of the Laboratory for Artificial Brain Systems, at the Frontier Research Program RIKEN (Japan), in the Brain Information Processing Group, and since 1999, he has been Head of the Laboratory for Advanced Brain Signal Processing, Brain Science Institute RIKEN, Saitama, Japan. His current research interests include biomedical signal and image processing (especially blind signal/image processing), neural networks and their applications, learning theory and robust algorithms, generalization and extensions of independent and principal component analysis, optimization problems and nonlinear circuits and systems theory and their applications.

Dr. Cichocki is a member of several international Scientific Committees and the Associated Editor of the IEEE TRANSACTION ON NEURAL NETWORKS (since January 1998).

Shun-ichi Amari (M'71–SM'92–F'94–LF'06) was born in Tokyo, Japan, on January 3, 1936. He received the Dr. Eng. degree majoring in mathematical engineering from the Graduate School of the University of Tokyo, Tokyo, Japan, in 1963.

He worked as an Associate Professor at Kyushu University and the University of Tokyo, and then a Full Professor at the University of Tokyo, and is now Professor-Emeritus. He serves now as Director of RIKEN Brain Science Institute, Saitama, Japan. He has been engaged in research in wide areas of mathematical engineering, such as topological network theory, differential geometry of continuum mechanics, pattern recognition, and information sciences. In particular, he has devoted himself to mathematical foundations of neural network theory, including statistical neurodynamics, dynamical theory of neural fields, associative memory, self-organization, and general learning theory. Another main subject of his research is information geometry initiated by himself, which applies modern differential geometry to statistical inference, information theory, control theory, stochastic reasoning, and neural networks, providing a new powerful method of information sciences and probability theory.

Dr. Amari is President of Institute of Electronics, Information and Communication Engineers, Japan, and Past President of International Neural Networks Society. He received the Emanuel A. Piore Award and the Neural Networks Pioneer Award from the IEEE, the Japan Academy Award, the C&C award, among many others.

Shengli Xie (M'01–SM'02) was born in Hubei Province, China, in 1958. He received the M.S. degree in mathematics from Central China Normal University, Wuhan, China, in 1992 and the Ph.D. degree in control theory and applications from South China University of Technology, Guangzhou, China, in 1997.

Currently, he is a Full Professor at South China University of Technology, and a Vice Head of Institute of Automation and Radio Engineering. He is the author or coauthor of two books and more than 70 scientific papers in journals and conference proceedings. His research interests include automatic control and blind signal processing.

Cuntai Guan (S'91–M'97–SM'03) received the Ph.D. degree in electrical and electronic engineering from Southeast University, China, in 1993.

From 1993 to 1994, he was at the Southeast University, where he worked on speech vocoder, speech recognition, and text-to-speech. During 1995, he was a Visiting Scientist at the Centre de Recherche en Informatique de Nancy (CRIN)/Centre National de la Recherche Scientifique (CNRS)-Institut National de Recherche en Informatique et en Automatique (INRIA), France, where he was working on key word spotting. From 1996 to 1997, he was with the City University of Hong Kong, Hong Kong, where he was developing robust speech recognition under noisy environment. From 1997 to 1999, he was with the Kent Ridge Digital Laboratories, Singapore, where he was working on multilingual, large vocabulary, continuous speech recognition. He was a Research Manager and the R&D Director for five years in industries, focusing on the development of spoken dialogue technologies. In 2003, he established the Brain–Computer Interface Group at Institute for Infocomm Research, Agency for Science, Technology and Research (A*STAR), Singapore, where he is currently a Senior Scientist and a Program Manager. His current research interests include brain–computer interface, neural signal processing, machine learning, pattern classification, and statistical signal processing, with applications to assistive device, rehabilitation, and health monitoring.