# Probability Estimation for Recoverability Analysis of Blind Source Separation Based on Sparse Representation

Yuanqing Li, Shun-Ichi Amari, *Fellow, IEEE*, Andrzej Cichocki, *Member, IEEE*, and Cuntai Guan, *Senior Member, IEEE*

***Abstract*—An important application of sparse representation is underdetermined blind source separation (BSS), where the number of sources is greater than the number of observations. Within the stochastic framework, this paper discusses recoverability of underdetermined BSS based on a two-stage sparse representation approach. The two-stage approach is effective when the source matrix is sufficiently sparse. The first stage of the two-stage approach is to estimate the mixing matrix, and the second is to estimate the source matrix by minimizing the 1-norms of the source vectors subject to some constraints. After estimating the mixing matrix and fixing the number of nonzero entries of a source vector, we estimate the recoverability probability (i.e., the probability that the source vector can be recovered). A general case is then considered where the number of nonzero entries of the source vector is fixed and the mixing matrix is drawn from a specific probability distribution. The corresponding probability estimate on recoverability is also obtained. Based on this result, we further estimate the recoverability probability when the sources are also drawn from a distribution (e.g., Laplacian distribution). These probability estimates not only reflect the relationship between the recoverability and sparseness of sources, but also indicate the overall performance and confidence of the two-stage sparse representation approach for solving BSS problems. Several simulation results have demonstrated the validity of the probability estimation approach.**

***Index Terms*—Blind source separation (BSS), linear programming, probability estimation, recoverability, sparse representation.**

## I. Introduction

SPARSE representation, or sparse coding, of signals has received a great deal of attention in recent years (e.g., [1]–[10], etc). An important application of the sparse representation is in underdetermined blind source separation (BSS), where the number of sources is greater than the number of observations. Until now, the independent component analysis (ICA) approach has been commonly used to solve BSS problems. However, generally the ICA approach cannot recover all sources in the underdetermined case [12]–[15]. Moreover, the sparse representation approach can handle the following two cases, which are difficult to deal with by using the standard ICA approach: 1) the sources are dependent; and 2) the sources are nonstationary.

Several algorithms based on sparse representation have been developed for BSS. For instance, the mixing matrix and sources were estimated using the maximum posterior approach and the maximum-likelihood approach in [17]–[19]. A variational expectation maximization algorithm for sparse representation was proposed for underdetermined BSS [20], and a two-stage clustering-then-$l^1$-optimization approach was proposed for underdetermined BSS in which the mixing matrix and the sources were estimated separately [16], [18]. In addition, several studies analyzed perturbations of sparse signal representation in the presence of noise [21]–[23].

Recently, in [21], we analyzed the two-stage clustering-then-$l^1$-optimization approach for sparse representation and its application to BSS. The uniqueness of the $l^1$ norm solution and its robustness to additive noise were discussed. The equivalence of the $l^1$-norm solution and the $l^0$-norm solution was also discussed within the context of a probabilistic framework. These results were then used in a recoverability analysis to BSS.

In this paper, a general case is considered in which the mixing matrix and the source matrix are taken randomly. We estimate the probability of recoverability when the two-stage sparse representation approach is used for underdetermined BSS.

First, we present the model and explain the two-stage sparse representation approach. We consider the following model neglecting noise:

$$\boldsymbol{X} = \boldsymbol{A}\boldsymbol{S} \tag{1}$$

where the unknown mixing matrix $\boldsymbol{A} \in R^{n\times m}$, the matrix $\boldsymbol{S} = [\boldsymbol{s}(1), \dots, \boldsymbol{s}(K)] \in R^{m\times K}$ contains $m$ unknown sources, and the only observable $\boldsymbol{X} = [\boldsymbol{x}(1), \dots, \boldsymbol{x}(K)] \in R^{n\times K}$ is a data matrix containing $n$ mixtures of the sources. In general, the number of sources, $m$, is also unknown. In this paper, we focus on the underdetermined case in which $n < m$.

In the above model, a column $\boldsymbol{s}(k) \in R^m$ $(k = 1, \dots, K)$ of $\boldsymbol{S}$ is called a source vector. The task of BSS is to recover the sources using only the observable matrix $\boldsymbol{X}$. In the two-stage sparse representation approach, the mixing matrix $\boldsymbol{A}$ is estimated in the first stage, and the source matrix $\boldsymbol{S}$ is estimated

Manuscript receivedApril 14, 2005; revised March 3, 2006.

Y. Li and C. Guan are with the Institute for Infocomm Research, Singapore 119613.

S.-I. Amari is with the Laboratory for Mathematical Neuroscience, RIKEN Brain Science Institute, Wako shi, Saitama, 3510198, Japan.

A. Cichocki is with the Laboratory for Advanced Brain Signal Processing RIKEN Brain Science Institute Wako shi, Saitama, 3510198, Japan.

Communicated by X. Wang, Associate Editor for Detection and Estimation.

Digital Object Identifier 10.1109/TIT.2006.876348

in the second stage by solving the following set of $K$ optimization problems:

$$\min \sum_{i=1}^{m} |s_i(j)|, \quad \text{s. t. } \boldsymbol{As}(j) = \boldsymbol{x}(j) \tag{2}$$

where $j = 1, \ldots, K$.

*Remark 1:* As will be seen in the following sections, the sparsity of source components plays a key role in the two-stage approach. The high sparsity of sources can guarantee a high probability that the sources can be recovered. Many sources (e.g., speech and image sources) are not sparse in the time domain but are sparse in the time–frequency domain if a suitable linear transformation is applied. Thus, if necessary, we can consider the problem in the time–frequency domain rather than in the time domain [21]. All discussions in this paper can be applied in the time domain, the frequency domain, and the time–frequency domain.

When the mixing matrix $\boldsymbol{A}$ is correctly (or sufficiently accurately) estimated, we can then discuss the recoverability problem. We now rephrase the problem as a question. How is it possible for the 1-norm solution of (2) to be equal to the true source vector?

For simplification, discussion in the following sections will be based on the following optimization problem instead of (2), which can be seen as a representative of the $K$ optimization problems in (2):

$$(P_1) \min \sum_{i=1}^{m} |s_i|, \quad \text{s. t. } \boldsymbol{As} = \boldsymbol{x}^*$$

where $\boldsymbol{x}^* = \boldsymbol{As}^*$, $\boldsymbol{s}^* \in R^m$, is a true source vector.

In this paper, let $\boldsymbol{s}_1$ denote a solution of $(P_1)$. From the discussion in [21], the 1-norm solution of $(P_1)$ is unique with probability one. In this paper, we mainly consider the probability that the 1-norm solution is equal to the source vector. If the mixing matrix $\boldsymbol{A}$ is given or already estimated, and the number $l$ of nonzero entries of $\boldsymbol{s}^*$ is fixed (i.e., $\|\boldsymbol{s}^*\|_0 = l$, $0 \le l \le m$), the probability can then be denoted as the conditional probability $P(\boldsymbol{s}_1 = \boldsymbol{s}^*; \|\boldsymbol{s}^*\|_0 = l, \boldsymbol{A})$.

For more general cases, where all entries of $\boldsymbol{A} \in R^{n \times m}$ are also drawn from a distribution (e.g., a uniform distribution valued in an interval), the recoverability probability then is determined as a function of the sensor number $n$, the source number $m$, and the number $l$ of nonzero entries of $\boldsymbol{s}^*$. Hence, we denote

$$P(n, m, l) = P(\boldsymbol{s}_1 = \boldsymbol{s}^*; \|\boldsymbol{s}^*\|_0 = l). \tag{3}$$

In this paper, we focus on how to evaluate the probability in (3). After obtaining the estimation of $P(n, m, l)$, we can estimate the probability $P(\boldsymbol{s}_1 = \boldsymbol{s}^*)$ when the number of nonzero entries of the source vector is not fixed but the sources are drawn from a distribution (e.g., Laplacian distribution).

The remainder of this paper is organized as follows: Section II presents some preliminary results when the mixing matrix is estimated or given. Section III discusses the probability estimation of $P(n, m, l)$ in (3). Section IV includes several simulation examples, and discusses the probability estimation when the mixing matrix and source matrix are drawn from several given distributions. Finally, in Section V, we review the results obtained in this paper.

## II. Preliminaries

As mentioned in the previous section, the first stage of the two-stage BSS approach is to estimate the mixing matrix. The source vectors are then estimated by solving the linear programming problem (2). In this section, we present some preliminary results. Several recoverability results are obtained when the mixing matrix is given or estimated. Specifically, the probabilities $P(\boldsymbol{s}_1 = \boldsymbol{s}^*; \|\boldsymbol{s}^*\|_0 = l, \boldsymbol{A})$ $(l = 1, \ldots, m)$ are estimated.

Considering the linear programming problem $(P_1)$, we denote $G$ as the index set of nonzero entries of $\boldsymbol{s}^*$, i.e., $G = \{i_1, \ldots, i_l | s^*_{i_j} \neq 0, j = 1, \ldots, l\}$, and $F$ as the set of all the subsets of $G$ except the null set. The cardinality of $F$ is $2^l - 1$. In the following theorem, we present two sets of necessary and sufficient conditions on the recoverability of $\boldsymbol{s}^*$, in which the first one is an extension of a result in [11].

*Theorem 1:*

1. $\boldsymbol{s}^* = \boldsymbol{s}_1$, if and only if, $\forall I \in F$, the optimal value of the objective function of the following optimization problem (4) is less than $\frac{1}{2}$, provided that it is solvable:

$$\begin{aligned} &\max \sum_{k \in I} |\delta_k|, \quad \text{s. t.} \\ &\boldsymbol{A\delta} = 0, \quad \|\boldsymbol{\delta}\|_1 = 1 \\ &\delta_k s^*_k > 0 \text{ for } k \in I \\ &\delta_k s^*_k \le 0 \text{ for } k \in G \setminus I. \end{aligned} \tag{4}$$

2. $\boldsymbol{s}^* = \boldsymbol{s}_1$, if and only if the optimal value of the objective function of the following optimization problem is less than $\frac{1}{2}$:

$$\begin{aligned} &\max \sum_{k=1}^{m} [\text{sign}(s^*_k)\delta_k]_+, \quad \text{s. t.} \\ &\boldsymbol{A\delta} = 0, \quad \|\boldsymbol{\delta}\|_1 = 1 \end{aligned} \tag{5}$$

   where $\forall y \in R, [y]_+ = \max(y, 0)$.

The proof of this theorem is given in Appendix I.

Note that it is not difficult to transform the optimization problem (4) into a standard linear programming problem (see Appendix II).

*Remarks 2:*

1. Although the 0-norm solution is the sparsest, as stated in [21], it is not easy to directly find such a solution by minimizing the 0-norm. The reasons are: a) the 0-norm solution is generally not unique if the number of its nonzero entries is $n$; b) there is no efficient algorithm to directly find the 0-norm solution; c) the 0-norm solution is not robust to noise. However, the 1-norm solution is not only sparse, but also advantageous over the 0-norm solution in these three aspects. Furthermore, it is also feasible to find the sparsest solution (the 0-norm solution) by minimizing the 1-norm.
2. In many references on basis pursuit, e.g., [11], the equivalence between the 0-norm and 1-norm solutions has been

studied. In Theorem 1, if we assume the true source vector $\boldsymbol{s}^*$ be the 0-norm solution (the sparsest solution), then the necessary and sufficient conditions for source recovery also hold for the equivalence between the 0-norm and 1-norm solutions.

3. Recently, the analysis of equivalence between between the 0-norm and 1-norm solutions under a probabilistic framework was reported, e.g., [25], [26], [28].

*Remark 3:* From Theorem 1, we can find that for a given mixing matrix $\boldsymbol{A}$, the recoverability of the source column vector $\boldsymbol{s}^*$ obtained by solving the problem $(P_1)$ depends on the index set $G$ and signs of its nonzero entries, but not on the magnitudes of its entries. This observation has been done in earlier papers, e.g., [24]. When estimating the mixing matrix, we often normalize the estimated columns of the mixing matrix to length one. Obviously, this will change the amplitudes of the recovered sources. However, the normalization procedure does not affect the recoverability of the sources. Note that, in this paper, if an estimated source is equal to a true source up to scaling, then we say that the true source can be recovered.

The following is the uniqueness theorem for the solution of $(P_1)$.

*Theorem 2:* The solution $\boldsymbol{s}_1$ of $(P_1)$ is unique if and only if, $\forall I \in F_1$, the optimal value of the objective function of the following optimization problem (6) is less than $\frac{1}{2}$, provided that it is solvable:

$$\begin{aligned} &\max \sum_{k\in I} |\delta_k|, \quad \text{s. t.} \\ &\boldsymbol{A\delta} = 0, \quad \|\boldsymbol{\delta}\|_1 = 1 \\ &\delta_k s_{1k} > 0, \ \text{for } k \in I \\ &\delta_k s_{1k} \le 0, \ \text{for } k \in G_1 \setminus I \end{aligned} \tag{6}$$

where $G_1$ is the index set of nonzero entries of $\boldsymbol{s}_1$, and $F_1$ is the set of all the subsets of $G_1$ except the null set. Furthermore, $s_{1k}$ is the $k-$th entry of $\boldsymbol{s}_1$.

The proof is straightforward from Theorem 1.

We now estimate the conditional probability $P(\boldsymbol{s}^* = \boldsymbol{s}_1; \|\boldsymbol{s}^*\|_0 = l, \boldsymbol{A})$. Note that there are $C_m^l (= \frac{m!}{l!(m-l)!})$ index subsets of $G_0$ with cardinality $l$, where $G_0 = \{1, 2, \ldots, m\}$. We denote these subsets as $G_k^l$, $k = 1, \ldots, C_m^l$.

In this section, we need the following assumption with respect to the source vector.

*Assumption 1:* All nonzero entries of the source vector $\boldsymbol{s}^*$ take either positive or negative sign with equal probability. The index set $G$ of its $l$ nonzero entries can be one of the $C_m^l$ index sets $G_k^l$, where $k = 1, \ldots, C_m^l$ with equal probability, i.e.,

$$P\left(G = G_k^l\right) = \frac{1}{C_m^l}, \quad \text{for } k = 1, \ldots, C_m^l.$$

Obviously, we first have

$$P(\boldsymbol{s}^* = \boldsymbol{s}_1; \|\boldsymbol{s}^*\|_0 = 0, \boldsymbol{A}) = P(\boldsymbol{s}^* = \boldsymbol{s}_1; \|\boldsymbol{s}^*\|_0 = 1, \boldsymbol{A}) = 1$$

where the second equality is obtained from Theorem 1. Following the linear programming theory, we know that $\boldsymbol{s}_1$ has at most $n$ nonzero entries. Hence, the true source vector $\boldsymbol{s}^*$ is not equal to $\boldsymbol{s}_1$ when $n < \|\boldsymbol{s}_1\|_0 \le m$, that is,

$$P(\boldsymbol{s}^* = \boldsymbol{s}_1; \|\boldsymbol{s}^*\|_0 = l, \boldsymbol{A}) = 0$$

for $l = n+1, \ldots, m$.

Suppose that $\boldsymbol{s}^*$ has $l$ $(2 \le l \le n)$ nonzero entries, and the index set of its nonzero entries is $G$. Using $\boldsymbol{s}^*$, we can define a sign column vector $\boldsymbol{t} = sign(\boldsymbol{s}^*) \in R^m$.

Theorem 1 and Remark 3 indicate that the recoverability of a source vector $\boldsymbol{s}^*$ is only related to the index set and signs of its nonzero entries, but not related to the amplitudes of its nonzero entries. Thus, the recoverability of $\boldsymbol{s}^*$ is equivalent to that of $\boldsymbol{t}$.

For a given index set $G_k^l$ and a sign column vector $\boldsymbol{t}$, whose nonzero entries index set is $G_k^l$, if for any nonnull subset $I \subseteq G_k^l$, the optimal value of (4) is less than $\frac{1}{2}$, then $\boldsymbol{t}$ can be recovered by solving the linear programming problem $(P_1)$ when $\boldsymbol{s}^*$ is replaced by $\boldsymbol{t}$.

Noting that there are $2^l$ sign column vectors whose nonzero entries index set is $G_k^l$, suppose that there are $q_k^l$ sign column vectors that can be recovered, then $\frac{q_k^l}{2^l}$ is the probability that a source vector $\boldsymbol{s}^*$, with its nonzero entries index set being $G_k^l$, can be recovered by solving $(P_1)$. Therefore, we have

$$P(\boldsymbol{s}^* = \boldsymbol{s}_1; \|\boldsymbol{s}^*\|_0 = l, A) = \sum_{k=1}^{C_m^l} \frac{q_k^l}{2^l} \frac{1}{C_m^l} = \frac{1}{2^l C_m^l} \sum_{k=1}^{C_m^l} q_k^l \tag{7}$$

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

*Remarks 4:*

1. If the mixing matrix $\boldsymbol{A}$ is given or estimated, the key number $q_k^l$ in (7) can be determined by checking whether the necessary and sufficient condition (4) is satisfied for all the $2^l$ sign-column vectors related to the index set $G_k^l$.
2. Generally, the estimation approach in (7) is suitable for relatively small scale problem (e.g., $m < 15$). The computational complexity of (7) to estimate $P(\boldsymbol{s}^* = \boldsymbol{s}_l; \|\boldsymbol{s}^*\|_0 = l, A)$ will grow exponentially with $m$. However, when $m$ is large, we can estimate $P(\boldsymbol{s}^* = \boldsymbol{s}_l; \|\boldsymbol{s}^*\|_0 = l, A)$ by simulations similarly as in Example 3. The method is to randomly extract some sign vectors from $C_m^l 2^l$ ones and check whether they can be recovered or not. This way, we can estimate the recoverability probability. Since it is not necessary to check all sign vectors, the computational complexity will not increase dramatically.

## III. Probability of Recoverability Under a Random Mixing Matrix

In the preceding section, we presented the estimate of conditional probability $P(\boldsymbol{s}^* = \boldsymbol{s}_1;\ \|\boldsymbol{s}^*\|_0 = l, A)$ under a given or estimated mixing matrix $\boldsymbol{A}$. In this section, we consider a general case in which the mixing matrix is randomly generated, and estimate the probability $P(n, m, l)$ in (3). We assume that all the entries of the mixing matrix are drawn from the uniform distribution in $[-1, 1]$ independently.

First, consider two standard linear programming problems

$$\begin{aligned} &\min f(\boldsymbol{s}), \quad \text{s. t.} \\ &\boldsymbol{As} = \boldsymbol{x}, \quad \boldsymbol{s} \ge \boldsymbol{0} \end{aligned} \tag{8}$$

$$\min f(\boldsymbol{s}), \quad \text{s. t.}$$
$$(\boldsymbol{A}+\Delta\boldsymbol{A})\boldsymbol{s}=\boldsymbol{x}, \quad \boldsymbol{s}\geq \boldsymbol{0} \tag{9}$$

where $f(\cdot)$ is a linear function, $\boldsymbol{A}\in R^{n\times m}$, $\boldsymbol{x}\in R^n$, $n<m$. We suppose that problem (8) has an optimal solution denoted as $\bar{\boldsymbol{s}}$.

We have the following lemma on the linear programming problems (8) and (9).

*Lemma 1:* For liner programming problems (8) and (9), if the optimal solution $\bar{\boldsymbol{s}}$ of (8) has $n$ nonzero entries, and $||\Delta\boldsymbol{A}||$ is sufficiently small, then the linear programming problem (9) is feasible, and

$$\lim_{||\Delta \boldsymbol{A}||\to 0}||\bar{\boldsymbol{s}}-\bar{\boldsymbol{s}}_{\Delta\boldsymbol{A}}||=0, \quad \lim_{||\Delta \boldsymbol{A}||\to 0}|f(\bar{\boldsymbol{s}})-f(\bar{\boldsymbol{s}}_{\Delta\boldsymbol{A}})|=0 \tag{10}$$

where $\bar{\boldsymbol{s}}_{\Delta\boldsymbol{A}}$ denotes the optimal solution of (9).

The proof of Lemma 1 can be seen in Appendix III.

For linear programming problem (8), we have $||\bar{\boldsymbol{s}}||_0\leq n$ [21]. If $\boldsymbol{A}\in R^{n\times m}$ and $\boldsymbol{x}\in R^n$ are arbitrarily taken, then generally at least $n$ columns of $\boldsymbol{A}$ are necessary to linearly represent $\boldsymbol{x}$, that is, $\bar{\boldsymbol{s}}$ has $n$ nonzero entries with probability one. Thus, in reality, the condition on the number of nonzero entries of $\bar{\boldsymbol{s}}$ in Lemma 1 is easy to satisfy.

We now present a theorem on the conditional probability $P(\boldsymbol{s}^*=\boldsymbol{s}_1;||\boldsymbol{s}^*||_0=l,\boldsymbol{A})$.

*Theorem 3:* $p(\boldsymbol{s}^*=\boldsymbol{s}_1;|\boldsymbol{s}^*|_0=l,\boldsymbol{A})$ is continuous with respect to $\boldsymbol{A}$ except on a zero-measure set of $\boldsymbol{A}$.

The proof of Theorem 3 can be seen in Appendix IV.

From Theorem 3, $P\{\boldsymbol{s}^*=\boldsymbol{s}_1;||\boldsymbol{s}^*||_0=l,\boldsymbol{A}\}$ is continuous with respect to $\boldsymbol{A}$ almost everywhere, probability $P(n,m,l)$ can be expressed by the following integral:

$$P(n,m,l)=\int_{||A||_\infty\leq 1}P(\boldsymbol{s}^*=\boldsymbol{s}_1;||\boldsymbol{s}^*||_0=l,\boldsymbol{A})P(\boldsymbol{A})d\boldsymbol{A} \tag{11}$$

where $||A||_\infty=\max\{|a_{ij}|,i=1,\dots,n,j=1,\dots,m\}$.

Although conditional probability $P\{\boldsymbol{s}^*=\boldsymbol{s}_1;||\boldsymbol{s}^*||_0=l,\boldsymbol{A}\}$ can be estimated for each $\boldsymbol{A}$ using (7), it is difficult to obtain an explicit result of (11). In the following, we will give a good approximation of the integral in (11). The validity of the approximation has been demonstrated by our simulations.

For a mixing matrix $\boldsymbol{A}$ and a sign vector $\boldsymbol{t}$ with $l$ nonzero entries, if the optimal value of optimization problem (5) is less than $\frac{1}{2}$, then $\boldsymbol{t}$ can be recovered. If so, $(\boldsymbol{A},\boldsymbol{t})$ is said to be recoverable. Let $M(\boldsymbol{A},\boldsymbol{t})$ denote the optimal value of the objective function in optimization problem (5) with $\boldsymbol{s}^*$ replaced by $\boldsymbol{t}$ and $P(\boldsymbol{t})$ denote the probability that the source vector $\boldsymbol{t}$ is equal to the corresponding 1-norm solution under a random mixing matrix $\boldsymbol{A}$.

For deriving our approximation to the probability in (11), we now construct the table at the bottom of the page.

In the table, $\boldsymbol{t}(1),\dots,\boldsymbol{t}(N)$ are all $N\ \left(=2^l*C_m^l\right)$ sign vectors with $l$ nonzero entries, $\boldsymbol{A}_1,\dots,\boldsymbol{A}_N$ are $N$ independent and identically distributed (i.i.d.) samples of $\boldsymbol{A}$ drawn from the uniform distribution.

Let $I(\boldsymbol{A},\boldsymbol{t})$ be the indicator function such that $(\boldsymbol{A},\boldsymbol{t})$ is recoverable, that is, $I(\boldsymbol{A},\boldsymbol{t})=1$ when $M(\boldsymbol{A},\boldsymbol{t})<1/2$ and 0 otherwise. Hence,

$$P(I(\boldsymbol{A},\boldsymbol{t})=1)=P(\boldsymbol{t}), \quad E(I(\boldsymbol{A},\boldsymbol{t}))=P(\boldsymbol{t}). \tag{12}$$

Then, for any sign vector $\boldsymbol{t}$, the recoverability probability $P(\boldsymbol{t})$ can be approximated by

$$P(\boldsymbol{t})\approx\frac{1}{N}\sum_{i=1}^{N}I\left(\boldsymbol{A}_i,\boldsymbol{t}\right). \tag{13}$$

When $N$ is large, the law of large numbers guarantees the above approximation.

According to the probability estimation in (7) and the definition of $I(\boldsymbol{A}_i,\boldsymbol{t}(q))$, we have

$$P(\boldsymbol{s}^*=\boldsymbol{s}_1;||\boldsymbol{s}^*||_0=l,\boldsymbol{A}_i)=\frac{\sum_{k=1}^{N}I(\boldsymbol{A}_i,\boldsymbol{t}(k))}{N}. \tag{14}$$

*Theorem 4:* Suppose that all entries of $\boldsymbol{A}$ are drawn from the uniform distribution in $[-1,1]$, then $N$ sign vectors $\boldsymbol{t}(1),\dots,\ \boldsymbol{t}(N)$ (with $l$ nonzero entries) can be recovered by solving linear programming problem $(P_1)$ with equal probabilities, i.e.,

$$P(\boldsymbol{t}(1))=P(\boldsymbol{t}(2))=\cdots=P(\boldsymbol{t}(N)). \tag{15}$$

The proof of Theorem 4 can be seen in Appendix V.

*Proposition 1:* For the sign vectors $\boldsymbol{t}(1),\dots,\boldsymbol{t}(N)$, most of event pairs $\left(M(\boldsymbol{A},\boldsymbol{t}(q_1))<\frac{1}{2}\right)$ and $\left(M(\boldsymbol{A},\boldsymbol{t}(q_2))<\frac{1}{2}\right)$ are close to being uncorrelated especially when $l$ is much less than $n$ or close to $n$, where $q_1,q_2=1,\dots,N$, $q_1\neq q_2$, $\boldsymbol{A}$ is randomly taken from a specific distribution.

| | $P(\boldsymbol{s}^*=\boldsymbol{s}_1;\|\boldsymbol{s}^*\|_0=l,\boldsymbol{A}_1)$ | $P(\boldsymbol{s}^*=\boldsymbol{s}_1;\|\boldsymbol{s}^*\|_0=l,\boldsymbol{A}_2)$ | $\dots$ | $P(\boldsymbol{s}^*=\boldsymbol{s}_1;\|\boldsymbol{s}^*\|_0=l,\boldsymbol{A}_N)$ |
|---|---|---|---|---|
| $P(\boldsymbol{t}(1))$ | $M(\boldsymbol{A}_1,\boldsymbol{t}(1))$ | $M(\boldsymbol{A}_2,\boldsymbol{t}(1))$ | $\dots$ | $M(\boldsymbol{A}_N,\boldsymbol{t}(1))$ |
| $P(\boldsymbol{t}(2)))$ | $M(\boldsymbol{A}_1,\boldsymbol{t}(2))$ | $M(\boldsymbol{A}_2,\boldsymbol{t}(2))$ | $\dots$ | $M(\boldsymbol{A}_N,\boldsymbol{t}(2))$ |
| $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ | $\vdots$ |
| $P(\boldsymbol{t}(N))$ | $M(\boldsymbol{A}_1,\boldsymbol{t}(N))$ | $M(\boldsymbol{A}_2,\boldsymbol{t}(N))$ | $\dots$ | $M(\boldsymbol{A}_N,\boldsymbol{t}(N))$ |

In Appendix VI, a sketch of proof of Proposition 1 can be seen. Additionally, here we show a simple example which may be useful for understanding Proposition 1. Suppose that $x_1,\ldots,x_L$ are $L$ i.i.d. random variables (say $N(0,1)$). Let us impose the constraint

$$\sum_{i=1}^{L}|x_i| = c,$$

where $c \geq 0$ is a constant. Then, the conditional probability $p(x_1,\ldots,x_L|c)$ is not independent. However, $\forall i,j \in \{1,\ldots,L,\ i \neq j,\ p(x_i,x_j|c)$ is asymptotically pairwise independent when $L$ tends to infinity.

From our simulations (Example 1), we find that if the number $l$ of nonzero entries is very small (close to 1) or very big (close to $n$), then any two of these events are almost uncorrelated. Otherwise, their correlation will increase. This is consistent with the explanation in the sketch of proof of Proposition 1.

In the following, we present a weak law of large numbers to a correlated case.

*Lemma 2:* Suppose that $x[i]$, $i = 1,\ldots,$ are a series of random variables, each of which having an identical distribution, and that most of random variable pairs $x[i]$ and $x[j]$ are uncorrelated. Then weak law of large numbers is satisfied for $\{x[i]\}$.

The proof of Lemma 2 can be seen in Appendix VII.

Consider random variables $I(\boldsymbol{A}_i,\boldsymbol{t}(q)), q = 1,\ldots,N$. From the above Proposition 1, and the definition of $I(\boldsymbol{A}_i,\boldsymbol{t}(q))$, we find most of variable pairs $I(\boldsymbol{A}_i,\boldsymbol{t}(q_1))$ and $I(\boldsymbol{A}_i,\boldsymbol{t}(q_2))$ are close to being uncorrelated. Since $\boldsymbol{A}_i$ is a random sample of $\boldsymbol{A}$, from Theorem 4, we have

$$P(I(\boldsymbol{A}_i,\boldsymbol{t}(1)) = 1) = \cdots = P(I(\boldsymbol{A}_i,\boldsymbol{t}(N)) = 1)$$

that is, $I(\boldsymbol{A}_i,\boldsymbol{t}(q))$ have an identical distribution.

Noting that $N$ is a large number, from (12) and Lemma 2, we have

$$P(\boldsymbol{t}(q)) = E(I(\boldsymbol{A}_j,\boldsymbol{t}(q))) \approx \frac{\sum\limits_{q=1}^{N} I(\boldsymbol{A}_j,\boldsymbol{t}(q))}{N}. \tag{16}$$

By (13) and (16)

$$\frac{\sum\limits_{i=1}^{N} I(\boldsymbol{A}_i,\boldsymbol{t}(q))}{N} \approx \frac{\sum\limits_{q=1}^{N} I(\boldsymbol{A}_j,\boldsymbol{t}(q))}{N}. \tag{17}$$

Furthermore, from (13), (14), and (17), we have

$$\begin{aligned} P(\boldsymbol{t}(q)) &\approx \frac{\sum\limits_{i=1}^{N} I(\boldsymbol{A}_i,\boldsymbol{t}(q))}{N} \\ &\approx \frac{\sum\limits_{k=1}^{N} I(\boldsymbol{A}_j,\boldsymbol{t}(k))}{N} \\ &= P(\boldsymbol{s}^* = \boldsymbol{s}_1; \|\boldsymbol{s}^*\|_0 = l, \boldsymbol{A}_j), \end{aligned} \tag{18}$$

where $q,j = 1,\ldots,N$.

From the proof of Theorem 3, we find that in a sufficiently small region of $\boldsymbol{A}$, the conditional probability $P(\boldsymbol{s}^* = \boldsymbol{s}_1; \|\boldsymbol{s}^*\|_0 = l, \boldsymbol{A})$ is a constant with respect to $\boldsymbol{A}$. Furthermore, from the expressions in (18), we have the following observation.

*Observation:* For most $\boldsymbol{A}, P(\boldsymbol{s}^* = \boldsymbol{s}_1; \|\boldsymbol{s}^*\|_0 = l, \boldsymbol{A})$ takes almost the same value (not depending on $\boldsymbol{A}$). That is,

$$\begin{aligned} P(n,m,l) &= \int_{\|A\|_\infty<1} P(\boldsymbol{s}^* = \boldsymbol{s}_1; \|\boldsymbol{s}^*\|_0 = l, \boldsymbol{A}) P(\boldsymbol{A}) d\boldsymbol{A} \\ &\approx P(\boldsymbol{s}^* = \boldsymbol{s}_1; \|\boldsymbol{s}^*\|_0 = l, \boldsymbol{A}_0) \\ &= \frac{1}{2^l C_m^l} \sum_{k=1}^{C_m^l} q_k^l \end{aligned} \tag{19}$$

where $\boldsymbol{A}_0$ is a random sample of $\boldsymbol{A}$, and the last equality is based on (7).

Besides the above theoretical justification, the observation is also confirmed by our simulations.

Since the events

$$\left(M(\boldsymbol{A},\boldsymbol{t}(1)) < \frac{1}{2}\right),\ldots,\left(M(\boldsymbol{A},\boldsymbol{t}(N)) < \frac{1}{2}\right)$$

are not strictly independent, and there exists a zero measure set of $\boldsymbol{A}$ where $P(\boldsymbol{s}^* = \boldsymbol{s}_1; \|\boldsymbol{s}^*\|_0 = l, \boldsymbol{A})$ is not continuous, we recommend using the following mean to approximate the probability $P(n,m,l)$:

$$P(n,m,l) \approx \frac{1}{k_0} \sum_{q=1}^{k_0} P(\boldsymbol{s}^* = \boldsymbol{s}_1; \|\boldsymbol{s}^*\|_0 = l, \boldsymbol{A}_q) \tag{20}$$

where $\boldsymbol{A}_1,\ldots,\boldsymbol{A}_{k_0}$ are $k_0$ random samples of $\boldsymbol{A}$. Here, because our estimation of integral (11) is not based on a sampling method, $k_0$ does not need to be large (5 in our simulation examples). Additionally, the probability $P(\boldsymbol{s}^* = \boldsymbol{s}_1; \|\boldsymbol{s}^*\|_0 = l, \boldsymbol{A}_q)$ can be calculated according to (7).

*Remark 5:* If all entries of $\boldsymbol{A}$ are drawn from a zero-mean, symmetrical distribution (e.g., a zero-mean Gaussian distribution), it is not difficult to find that the above analysis and conclusions still hold.

We have established the estimation of the probability $P(n,m,l)$. Next, we shall consider two cases where the sources are drawn from two different distributions, and derive the corresponding probability estimates on recoverability, where the mixing matrix is always taken randomly according to the uniform distribution in $[-1,1]$.

*Case a:* For any source vector $\boldsymbol{s}^* \in R^m$, suppose that the probabilities

$$P(s_k^* = 0) = \alpha, \quad P(s_k^* \neq 0) = 1-\alpha, \quad k = 1,\ldots,m \tag{21}$$

and that all nonzero entries of $\boldsymbol{s}^*$ have positive sign or negative sign with equal probability.

In Case a, the number of nonzero entries of source vectors are not fixed. We can find that the probability of $\boldsymbol{s}^*$ having $j$

nonzero entries is $C_m^j(1-\alpha)^j\alpha^{(m-j)}$. We have the probability estimation as follows:

$$\begin{aligned} pro(\alpha) &= P(\boldsymbol{s}^* = \boldsymbol{s}_1) \\ &= \sum_{j=0}^{m} P(\boldsymbol{s}^* = \boldsymbol{s}_1; \|\boldsymbol{s}^*\|_0 = j) P(\|\boldsymbol{s}^*\|_0 = j) \\ &= \sum_{j=0}^{m} C_m^j (1-\alpha)^j \alpha^{(m-j)} P\{\boldsymbol{s}^* = \boldsymbol{s}_1; \|\boldsymbol{s}^*\|_0 = j\} \\ &= \sum_{j=0}^{m} C_m^j (1-\alpha)^j \alpha^{(m-j)} P(n,m,j) \end{aligned} \tag{22}$$

where the probability estimates $P(n,m,j)$ $(j = 0,\ldots,m)$ can be calculated by (20).

*Case b:* Suppose that all entries of the source vector $\boldsymbol{s}^* \in R^m$ are drawn from a Laplacian distribution with probability density function $\frac{\lambda}{2}\exp(-\lambda|x|)$, where $\boldsymbol{s}^*$ has $m$ nonzero entries generally. Strictly speaking, $\boldsymbol{s}^*$ cannot be recovered by solving linear programming problem $(P_1)$, however, the 1-norm solution of $(P_1)$ can approximate the source vector as indicated in the following.

We first derive the theoretical probability estimate $P(\lambda) = P(\boldsymbol{s}^* = \boldsymbol{s}_1)$ when the mixing matrix $\boldsymbol{A}$ is drawn from a symmetrical uniform distribution.

Using $\boldsymbol{s}^*$, we define a column vector $\bar{\boldsymbol{s}}^* \in R^m$: $\bar{s}_k^* = 0$ if $|s_k^*| < \epsilon_0$, $\bar{s}_k^* = s_k^*$, if $|s_k^*| \geq \epsilon_0$. Here, $\epsilon_0$ is a small positive constant.

Consider the optimization problem $(P_1)$ with $\boldsymbol{x}^*$ replaced by $\bar{\boldsymbol{x}}^* = \boldsymbol{A}\bar{\boldsymbol{s}}^*$. Denote its $l^1$ norm solution as $\bar{\boldsymbol{s}}_1$ and denote the $l^1$ norm solution of $(P_1)$ corresponding to $\boldsymbol{s}^*$ as $\boldsymbol{s}_1$. It follows from the robustness analysis of the $l^1$ norm solution in [21] that, for a given small positive constant $\beta$, if $\epsilon_0$ is sufficiently small, then $\|\boldsymbol{s}_1 - \bar{\boldsymbol{s}}_1\|_2 < \beta$, i.e., $\boldsymbol{s}_1 \approx \bar{\boldsymbol{s}}_1$. Moreover, we have

$$\begin{aligned} \|\boldsymbol{s}^* - \boldsymbol{s}_1\|_2 &\leq \|\boldsymbol{s}^* - \bar{\boldsymbol{s}}^*\|_2 + \|\bar{\boldsymbol{s}}^* - \bar{\boldsymbol{s}}_1\|_2 + \|\bar{\boldsymbol{s}}_1 - \boldsymbol{s}_1\|_2 \\ &\leq \sqrt{m}\epsilon_0 + \beta + \|\bar{\boldsymbol{s}}^* - \bar{\boldsymbol{s}}_1\|_2. \end{aligned} \tag{23}$$

Note that $\epsilon_0$ and $\beta$ are small. If $\bar{\boldsymbol{s}}^*$ is equal to $\bar{\boldsymbol{s}}_1$, we can say that $\boldsymbol{s}^* \approx \boldsymbol{s}_1$, and *vice versa*. Hence, we have

$$pro(\lambda) = P(\|\boldsymbol{s}^* - \boldsymbol{s}_1\|_2 \leq \beta_0) \approx P(\bar{\boldsymbol{s}}^* = \bar{\boldsymbol{s}}_1), \tag{24}$$

where $\beta_0 = \sqrt{m}\epsilon_0 + \beta$.

Using (24), the probability $P(\boldsymbol{s}^* = \boldsymbol{s}_1)$ can be obtained by estimating $P(\bar{\boldsymbol{s}}^* = \bar{\boldsymbol{s}}_1)$. The estimation of $P(\bar{\boldsymbol{s}}^* = \bar{\boldsymbol{s}}_1)$ has been discussed in Case a. Since all entries of $\boldsymbol{s}^*$ are drawn from a Laplacian distribution, it can be found that $P(|s_k^*| < \epsilon_0) = 1 - \exp(-\lambda\epsilon_0)$, denoted as $\alpha_\lambda$. The probability that $\boldsymbol{s}^*$ has $j$ entries greater than $\epsilon_0$ (i.e., $\bar{\boldsymbol{s}}^*$ has $j$ nonzero entries) is $C_m^j(1-\alpha_\lambda)^j\alpha_\lambda^{(m-j)}$, i.e.,

$$P(\|\bar{\boldsymbol{s}}^*\|_0 = j) = C_m^j(1-\alpha_\lambda)^j\alpha_\lambda^{(m-j)}.$$

Using (24) and (22), we have

$$\begin{aligned} pro(\lambda) &\approx P(\bar{\boldsymbol{s}}^* = \bar{\boldsymbol{s}}_1) \\ &= \sum_{j=0}^{m} P(\bar{\boldsymbol{s}}^* = \bar{\boldsymbol{s}}_1; \|\bar{\boldsymbol{s}}^*\|_0 = j) P(\|\bar{\boldsymbol{s}}^*\|_0 = j) \\ &= \sum_{j=0}^{m} C_m^j (1-\alpha_\lambda)^j \alpha_\lambda^{(m-j)} P(n,m,j), \end{aligned} \tag{25}$$

where the probability estimates $P(n,m,j)$ $(j = 0,\ldots,m)$ can be calculated by (20).

In order to calculate the probability in (25), there is a parameter $\epsilon_0$ to be determined in advance. After the sensor number $(n)$ and source number $(m)$ are determined, the parameter can be set by simulation as indicated in simulation Example 5.

## IV. Simulation Examples

In this section, we present several simulations to demonstrate the conclusions and probability estimates in the previous section.

*Example 1:* In this example, we demonstrate the conclusion in Proposition 1.

Since $N$ is generally large, we will encounter heavy computation burden when we check whether the $N$ events

$$\left(M(\boldsymbol{A}, \boldsymbol{t}(1)) < \frac{1}{2}\right), \ldots, \left(M(\boldsymbol{A}, \boldsymbol{t}(N)) < \frac{1}{2}\right)$$

are independent or not. Here we do partial validation.

Set $n = 7$, $m = 9$, i.e., $\boldsymbol{A} \in R^{7\times 9}$. For each $l$ $(l = 1,\ldots,7)$, we randomly take 16 pairs of sign vectors $(\boldsymbol{t}(i_j), \boldsymbol{t}(k_j))$ $(j = 1,\ldots,16)$, and estimate the probabilities and joint probabilities

$$P(M\Big(\boldsymbol{A}, \boldsymbol{t}(i_j)) < \frac{1}{2}\Big), \quad P\Big(M(\boldsymbol{A}, \boldsymbol{t}(k_j)) < \frac{1}{2}\Big),$$

and

$$P\Big(M(\boldsymbol{A}, \boldsymbol{t}(i_j)) < \frac{1}{2}, M(\boldsymbol{A}, \boldsymbol{t}(k_j)) < \frac{1}{2}\Big).$$

We then calculate the error $P_e$

$$\begin{aligned} P_e(l,j) &= P\left(M(\boldsymbol{A}, \boldsymbol{t}(i_j)) < \frac{1}{2}, M(\boldsymbol{A}, \boldsymbol{t}(k_j)) < \frac{1}{2}\right) \\ &\quad - P\left(M(\boldsymbol{A}, \boldsymbol{t}(i_j)) < \frac{1}{2}\right) \\ &\quad \times P\left(M(\boldsymbol{A}, \boldsymbol{t}(k_j)) < \frac{1}{2}\right) \end{aligned} \tag{26}$$

where $l = 1,\ldots,7$, $j = 1,\ldots,16$.

To estimate the probabilities

$$P\left(M(\boldsymbol{A}, \boldsymbol{t}(i_j)) < \frac{1}{2}\right), \quad P\left(M(\boldsymbol{A}, \boldsymbol{t}(k_j)) < \frac{1}{2}\right),$$

and

$$P\left(M(\boldsymbol{A}, \boldsymbol{t}(i_j)) < \frac{1}{2}, M(\boldsymbol{A}, \boldsymbol{t}(k_j)) < \frac{1}{2}\right)$$

we randomly take 3000 matrices $\boldsymbol{A} \in R^{7\times 9}$ based on a uniform distribution in $[-1, 1]$. Suppose that the event

$\left(M(\boldsymbol{A},\boldsymbol{t}(i_j) < \frac{1}{2}\right)$ occurs $L_1$ times for the 3000 mixing matrices, then $\frac{L_1}{3000}$ is the estimate of $P\left(M(\boldsymbol{A},\boldsymbol{t}(i_j)) < \frac{1}{2}\right)$. The other two probabilities can be similarly estimated.

We found that for different $i$ and $j$, $Perror(l,i)$ is very close to $Perror(l,j)$ ($l = 1,\ldots,7$, $i,j = 1,\ldots,16$). Averaging $P_e(l,j)$ with respect to $j$, we obtain an average probability error vector denoted as $P_{me}$

$$P_{me} = [0, -0.000, -0.006, -0.038, -0.045, -0.034, -0.005]. \tag{27}$$

In this paper, we only present the average probability errors due to limited page space.

We find that all entries of the above two average probability error vectors $P_{me}$ are quite close to zero. This demonstrates the conclusion in Proposition 1.

However, from the average error vectors $P_{me}$, we can find that when $l$ is close to 1 or to 7, the errors are smaller than those when $l$ takes medium values. Note that smaller probability error implies higher independence as well as less correlation. This is in accordance with the analysis in the proof of Proposition 1 (see Appendix VI).

*Example 2:* In this simulation example, we demonstrate the conclusion derived from (18)) by simulation

$$P(\boldsymbol{s}_1 = \boldsymbol{s}^*; \|\boldsymbol{s}^*\|_0 = l, \boldsymbol{A}_1) \approx P(\boldsymbol{s}_1 = \boldsymbol{s}^*; \|\boldsymbol{s}^*\|_0 = l, \boldsymbol{A}_2) \tag{28}$$

where $\boldsymbol{A}_1$ and $\boldsymbol{A}_2$ are two randomly chosen mixing matrices. We would like to emphasize that there may exist some particular matrices such that (28) is not satisfied, however, the set of those matrices has measure zero.

Similarly to the previous example, we have $n = 7$, $m = 9$. For each $l$ ($l = 1,\ldots,7$), we take eight different mixing matrices according to the uniform distribution in $[-1,1]$. For each given mixing matrix, we take 3000 source vectors $\boldsymbol{s}^*$ of which the $l$ nonzero entries are drawn from the uniform distribution, and their indices are also random. We then estimate the eight probabilities $P(\boldsymbol{s}_1 = \boldsymbol{s}^*; \|\boldsymbol{s}^*\|_0 = l, \boldsymbol{A}_j)$ ($j = 1,\ldots,8$) as in Example 1.

For $l = 1,\ldots,7$, we calculate the mean probability

$$\frac{1}{8}\sum_{j=1}^{8} P(\boldsymbol{s}_1 = \boldsymbol{s}^*; \|\boldsymbol{s}^*\|_0 = l, \boldsymbol{A}_j)$$

denoted as $\Pr(l)$, we obtain a mean probability column vector

$$\Pr = [1,\ 0.9816,\ 0.9014,\ 0.7267, 0.4916,\ 0.2610,\ 0.0900]^T.$$

Next, for $l = 1,\ldots,7$ and $j = 1,\ldots,8$, we calculate the error $P(\boldsymbol{s}_1 = \boldsymbol{s}^*; \|\boldsymbol{s}^*\|_0 = l, \boldsymbol{A}_j) - \Pr(l)$ (denoted as $Error(l,j)$), and obtain the following probability error matrix, see (29) at the bottom of the page.

Considering the $l$th ($l = 1,\ldots,7$) row in the error matrix above, we find that $P(\boldsymbol{s}_1 = \boldsymbol{s}^*; \|\boldsymbol{s}^*\|_0 = l, \boldsymbol{A}_j) \approx \Pr(l)$, $j = 1,\ldots,8$. Hence, the conclusion expressed by (28) is confirmed by simulation.

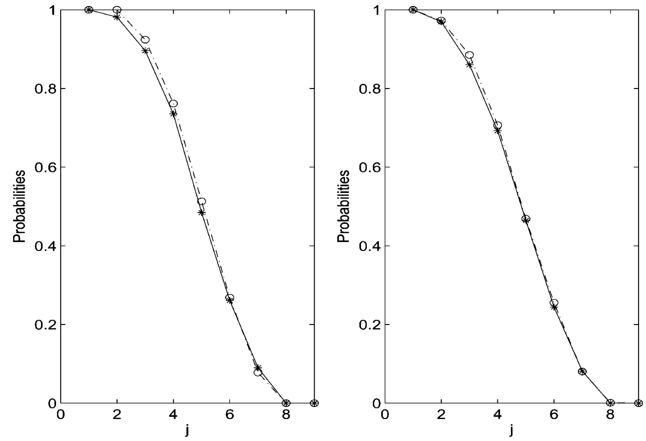


Fig. 1. Curves of estimated probabilities and true probabilities in Example 3. The solid curve with "$*$" in the left subplot depicts our probability estimates by (19), while the solid curve with "$*$" in the right subplot depicts our probability estimates by (20). The two dashed curves with "$\circ$" in the two subplots represent the true recoverability probabilities obtained by simulations.

*Example 3:* In this example, we demonstrate the validity of probability estimates obtained by (19) and (20) using simulations. Here, every mixing matrix $\boldsymbol{A} \in R^{7\times 9}$ is taken according to the uniform distribution in $[-1,1]$, and every source vector is nine-dimensional.

We first estimate the probabilities $P(7,9,l)$ ($l = 1,\ldots,9$) by simulations. For every $l$ ($l = 1,\ldots,9$), we take 3000 pairs of mixing matrices and source vectors. Note that each source vector has exactly $l$ nonzero entries drawn from a uniform distribution valued in the range $[-0.5, 0.5]$ with their indices also taken randomly. For each pair of source vector and mixing matrix, we solve the linear programming problem $(P_1)$ and check whether the 1-norm solution is equal to the source vector. Suppose that $n_l$ source vectors can be recovered, we obtain the ratio $\bar{p}_l = \frac{n_l}{3000}$ that reflects the true probability $P(7,9,l)$ on recoverability. All $\bar{p}_l, l = 1,\ldots,9$, are depicted by "$\circ$" and the dashed curve in the left subplot of Fig. 1.

Next, we take an mixing matrix $\boldsymbol{A} \in R^{7\times 9}$ as above. For every index set $G_k^j \subseteq G_0 (k = 1,\ldots,C_m^j,\ j = 2,\ldots 7)$, we first find the number $q_k^j$ of sign vectors that can be recovered by solving the linear programming problem $(P_1)$, and then calculate the probabilities $P(\boldsymbol{s}_1 = \boldsymbol{s}^*; \|\boldsymbol{s}^*\|_0 = l, \boldsymbol{A})$, $l = 2,\ldots,7$

$$Error = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.0184 & 0.0040 & -0.0155 & -0.0138 & -0.0060 & -0.0088 & 0.0034 & 0.0184 \\ 0.0312 & -0.0060 & -0.0117 & -0.0398 & -0.0155 & -0.0210 & 0.0055 & 0.0571 \\ 0.0226 & -0.0169 & -0.0025 & -0.0371 & -0.0110 & -0.0210 & -0.0029 & 0.0688 \\ 0.0145 & -0.0113 & 0.0031 & -0.0442 & -0.0026 & -0.0188 & -0.0051 & 0.0644 \\ 0.0145 & -0.0176 & -0.0053 & -0.0229 & -0.0096 & -0.0046 & 0.0016 & 0.0438 \\ 0.0061 & 0.0100 & -0.0044 & -0.0056 & 0.0044 & 0.0006 & -0.0156 & 0.0044 \end{bmatrix}. \tag{29}$$

using (7). Note that $P(\boldsymbol{s}_1 = \boldsymbol{s}^*; \|\boldsymbol{s}^*\|_0 = 1, \boldsymbol{A}) = 1$ and $P(\boldsymbol{s}_1 = \boldsymbol{s}^*; \|\boldsymbol{s}^*\|_0 = 8, \boldsymbol{A}) = P(\boldsymbol{s}_1 = \boldsymbol{s}^*; \|\boldsymbol{s}^*\|_0 = 9, \boldsymbol{A}) = 0$, we thus obtain nine probability estimates that are depicted by "$*$" and the solid curve in the left subplot of Fig. 1.

Fig. 1 shows that the two curves virtually overlap to some degree, which means that the probability $P(\boldsymbol{s}_1 = \boldsymbol{s}^*; \|\boldsymbol{s}^*\|_0 = l, \boldsymbol{A})$ calculated by (19) reflects the recoverability when the mixing matrix is also taken randomly.

Considering (20), we randomly take five mixing matrices denoted as $\boldsymbol{A}_1, \ldots, \boldsymbol{A}_5 \in R^{7\times 9}$, and calculate the probability estimates $P(\boldsymbol{s}_1 = \boldsymbol{s}^*; \|\boldsymbol{s}^*\|_0 = l, \boldsymbol{A}_j)$ $(j = 1, \ldots, 5, l = 1, \ldots, 9)$ using (7). We then calculate the average probability estimates $\frac{1}{5}\sum_{i=1}^{5} P(\boldsymbol{s}^* = \boldsymbol{s}_1; \|\boldsymbol{s}^*\|_0 = l, \boldsymbol{A}_j)$ $(l = 1, \ldots, 9)$. These nine probability estimates are represented by "$*$" and the solid curve in the right subplot of Fig. 1, while $\bar{p}_j, j = 1, \ldots, 9$ are represented by "$\circ$" and the dashed curve in the same subplot.

These two curves match each other very well, which indicates that the probability $P(\boldsymbol{s}^* = \boldsymbol{s}_1; \|\boldsymbol{s}^*\|_0 = j)$ calculated by (20) reflects the recoverability under condition that the number of nonzero entries of the source vector is fixed.

Comparing the four curves in Fig. 1, we can find that the probability estimated by (20) can better reflect the true probability than that estimated by (19).

*Example 4:* In this example, we consider the case in which the source vectors are drawn from the distribution (21), and confirm the probability estimate (22) by simulation.

For $\alpha_k = (j-1)*0.1$ $(k = 1, \ldots, 11)$, we calculate the probabilities $pro(\alpha_k)$ in (22) noting that $P(n, m, j)$ $(j = 0, \ldots, 9)$ have been obtained in Example 3.

For each $k$ $(k = 1, \ldots, 11)$, we take 3000 pairs of mixing matrices and source vectors. As in Example 3, all these mixing matrices are $7 \times 9$-dimensional and their entries are randomly valued in $[-1, 1]$. The 3000 source vectors are taken as follows:

$$\begin{aligned} s_q^* &= 0, \ if \ |\tilde{s}_q^*| \le \frac{\alpha_j}{2} \\ s_q^* &= \tilde{s}_q^*, \ if \ |\tilde{s}_q^*| > \frac{\alpha_j}{2} \end{aligned} \tag{30}$$

where $\tilde{\boldsymbol{s}}^* \in R^9$ is drawn from a uniform distribution valued in $[-0.5, 0.5]$. From (30), we can see that the probability is $\alpha_j$ that each entry of a source vector is equal to zero.

For each pair of source vector and mixing matrix, we solve the linear programming problem $(P_1)$ and check whether the $l^1$ norm solution is equal to the source vector. Suppose that $m_k$ source vectors can be recovered, thus, we obtain the ratio $\tilde{pro}(\alpha_k) = \frac{m_k}{3000}$, which reflects the true probability $P(\boldsymbol{s}_1 = \boldsymbol{s}^*)$ under the distribution parameter $\alpha_k$.

In Fig. 2, $pro(\alpha_k), k = 1, \ldots, 11$ are depicted by "$*$" and the solid curve, while $\tilde{pro}(\alpha_k), k = 1, \ldots, 11$, are depicted by "$\circ$" and the dashed curve. These two curves fit very well, virtually overlapping. Thus, if we know the distribution of all entries of the mixing matrix and the probability that each entry of a source vector is equal to zero, using (20) and (22), we can estimate the probability that the source can be recovered by solving $(P_1)$.

*Remark 6:* The probability parameter $\alpha$ in (21) can be seen as a sparsity index of sources. Bigger $\alpha$ implies higher sparsity of sources. From Fig. 2, we can see that the recoverability probability increases with $\alpha$. To obtain satisfactory recoverability, the sources must be sufficiently sparse. In many cases, since the sources are sparse in the time–frequency domain other than in the time domain, we can apply wavelet packets transformation to the observable mixtures and then perform BSS.

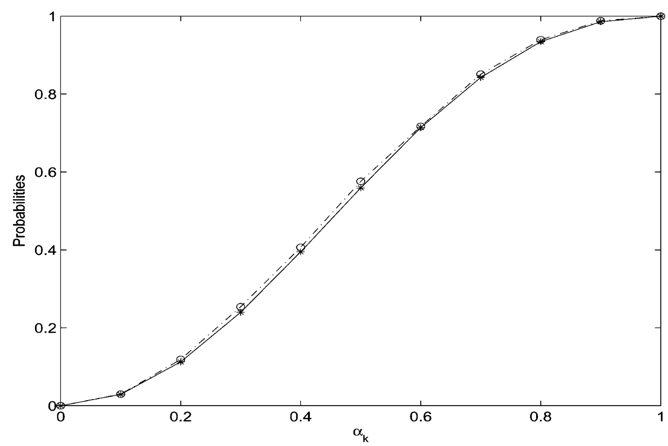


Fig. 2. Curves of estimated probabilities and true probabilities in Example 4, where the solid curve with "$*$" depicts our estimated probabilities, while the dashed curve with "$\circ$" represents the true recoverability probabilities. Note that the source vectors are drawn from the distribution in (30), the mixing matrices are drawn from a uniform distribution in $[-1, 1]$.

*Example 5:* In this example, we consider the case in which all entries of the source vector $\boldsymbol{s}^* \in R^9$ are drawn from a Laplacian distribution with probability density function $\frac{\lambda}{2}\exp(-\lambda|x|)$, and all entries of the $7 \times 9$-dimensional mixing matrix is drawn from a symmetrical uniform distribution in $[-1, 1]$. We present the simulation results to demonstrate the validity of the estimates obtained by (25).

In the following simulation, $\beta_0$ is set to be 0.055 and $\epsilon_0$ is set to be 0.023.

For 30 different Laplacian distributions with parameters $\lambda_k = 5k$ $(k = 1, \ldots, 30)$, we calculate the probabilities $pro(\lambda_k)$ using (25), where $\alpha_{\lambda_k} = 1 - exp(\lambda_k \epsilon_0)$ and $P(n, m, j)$ $(n = 7, m = 9, j = 0, \ldots, 9)$ were obtained in Example 3.

For every $k$ $(k = 1, \ldots, 30)$, 3000 pairs of mixing matrices and source vectors are taken. The source vectors are taken from a Laplacian distribution with parameter $\lambda_k$ and the 3000 mixing matrices $\boldsymbol{A} \in R^{7\times 9}$ are drawn from the same uniform distribution as in Example 3. The linear programming problem $(P_1)$ is solved for each pair of source vector $\boldsymbol{s}^*$ and mixing matrix. If the $l^1$ norm solution satisfies $\|\boldsymbol{s}_1 - \boldsymbol{s}^*\|_2 < \beta_0$, $\boldsymbol{s}^*$ can be recovered approximately. Suppose that there are $r_k$ source vectors recovered, we then have the ratio $\bar{pro}(\lambda_k) = \frac{r_k}{3000}$, which reflects the true probability $P(\boldsymbol{s}^* = \boldsymbol{s}_1)$.

Fig. 3 displays the curves of $pro(\lambda_k)$ (solid curve with "$*$") and $\bar{pro}(\lambda_k)$ (dashed curve with "$\circ$"). The two curves overlap remarkably well. From Fig. 3, we can see that as the parameter $\lambda$ in the Laplacian probability density function increases, the source vector becomes more and more sparse, and the recoverability probability increases.

From Fig. 3, we can see that when $\lambda_k$ is small or large (small $\lambda_k$ leads to dense source vectors, while large $\lambda_k$ leads to sparse source vectors), the two curves overlap better than when $\lambda_k$ is in medium. This is because our probability estimate (20) is based

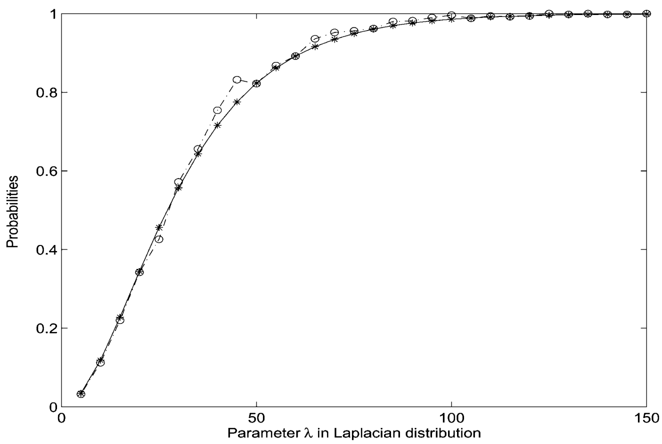


Fig. 3. Curves of (for) estimated probabilities and true probabilities in Example 5, where the solid curve with "$*$" depicts our estimated probabilities, while the dashed curve with "o" represents the true recoverability probabilities.

on Proposition 1, which can be well satisfied when source vectors are very sparse or very dense as indicated in Example 1.

Finally, with the above values of two parameters $\beta_0$ and $\epsilon_0$, the estimated probability using (25) well reflects the true recoverability probability for Laplacian source vectors. Thus, when $n = 7$, $m = 9$, and the mixing matrix is drawn from a symmetrical uniform distribution, the parameter $\epsilon_0$ can be set to be $0.023$, while the error between a recovered source vector and the original source vector is less than $0.055$.

## V. Concluding Remarks

When sparse representation is used in underdetermined BSS, recoverability problem should be dealt with as in ICA approach. However, this problem has not been completely solved. This paper continued discussion of this problem. The recoverability probability estimation was discussed when using a two-stage sparse-representation approach for BSS. First, we presented the conditional probability estimates of recoverability when the mixing matrix was given or estimated and the number of nonzero entries of a source vector was fixed. Next, we considered a more general case in which the mixing matrix was drawn from a symmetrical distribution (e.g., uniform distribution in $[-1, 1]$) and the number of nonzero entries was fixed. The recoverability probability estimate was then obtained. Using these estimated probabilities, we considered the recoverability probability estimation when the mixing matrix was drawn from a uniform distribution and source vectors were drawn from a distribution, e.g., Laplacian distribution, etc. Simulation results demonstrated that our probability estimates are very close to the true probabilities.

From the probability estimates in our simulations, we can find that the higher the sparsity of sources, the higher the recoverability probability. Additionally, these probability estimates show us the performance and confidence of sparse representation approach for solving underdetermined BSS problems. To obtain high recoverability probability, the sources must be sufficiently sparse. In many cases, this can be achieved by applying wavelet packets transformation to the observable mixtures. In practice, the level number of wavelet packets transformation can be determined by our probability estimate illustrated in our simulation Example 4. The probability estimates in this paper can provide us with some useful guide in solving underdetermined BSS problems.

## Appendix I
## Proof of Theorem 1

1. Necessity: Suppose that $\boldsymbol{s}^* = \boldsymbol{s}_1$; that is, $||\boldsymbol{s}^*||_1$ is the optimal value of $(P_1)$.

For a subset $I \in F$, when (4) is solvable, there is at least a flexible solution. For any flexible solution $\boldsymbol{\delta}$ of (4), it can be checked that $\boldsymbol{s}^* + t\boldsymbol{\delta}$ is a solution of the constraint equation of $(P_1)$, where $t < 0$ with sufficiently small absolute value. We have

$$\begin{aligned}||\boldsymbol{s}^* + t\boldsymbol{\delta}||_1 &= \sum_{k\in I}|s_k^* + t\delta_k| + \sum_{k\in G\setminus I}|s_k^* + t\delta_k| + \sum_{k\in G^c}|t\delta_k| \\ &= \sum_{k\in I}|s_k^*| - |t|\sum_{k\in I}|\delta_k| + \sum_{k\in G\setminus I}|s_k^*| \\ &\quad + |t|\sum_{k\in G\setminus I}|\delta_k| + |t|\sum_{k\in G^c}|\delta_k| \\ &= ||\boldsymbol{s}^*||_1 + |t|\left(\sum_{k\in I^c}|\delta_k| - \sum_{k\in I}|\delta_k|\right) \\ &> ||\boldsymbol{s}^*||_1. \end{aligned} \tag{31}$$

Thus,

$$\sum_{k\in I^c}|\delta_k| - \sum_{k\in I}|\delta_k| > 0. \tag{32}$$

It follows from $||\boldsymbol{\delta}||_1 = 1$ and (32) that $\sum_{k\in I}|\delta_k| < \frac{1}{2}$.

The necessity is proved.

Sufficiency: Suppose that $\boldsymbol{s}$ is a solution of the constraint equation in $(P_1)$, then $\boldsymbol{s}$ can be rewritten as

$$\boldsymbol{s} = \boldsymbol{s}^* + t^*\boldsymbol{\delta}, \tag{33}$$

where $\boldsymbol{\delta} = \frac{\boldsymbol{s}^* - \boldsymbol{s}}{||\boldsymbol{s}^* - \boldsymbol{s}||_1}$, $t^* = -||\boldsymbol{s}^* - \boldsymbol{s}||_1$.

Now we define an index set $I \in F$:

$$I = \{k | k = i_1, \ldots, i_l, \mathrm{sign}(s_k^*) = \mathrm{sign}(\delta_k)\}.$$

It can be checked easily that for the defined index set $I$, $\boldsymbol{\delta}$ is a flexible solution of (4). From the condition of the theorem, we have $\sum_{k\in I}|\delta_k| < \frac{1}{2}$. Furthermore

$$\begin{aligned}||\boldsymbol{s}^* + t\boldsymbol{\delta}||_1 &= \sum_{k\in I}|s_k^* + t\delta_k| + \sum_{k\in G\setminus I}|s_k^* + t\delta_k| + \sum_{k\in G^c}|t\delta_k| \\ &\geq \sum_{k\in I}|s_k^*| - |t|\sum_{k\in I}|\delta_k| + \sum_{k\in G\setminus I}|s_k^*| \\ &\quad + |t|\sum_{k\in G\setminus I}|\delta_k| + |t|\sum_{k\in G^c}|\delta_k| \\ &= ||\boldsymbol{s}^*||_1 + |t|\left(\sum_{k\in I^c}|\delta_k| - \sum_{k\in I}|\delta_k|\right) \\ &\geq ||\boldsymbol{s}^*||_1. \end{aligned} \tag{34}$$

Note that the equality in the last step of (34) holds when $t = 0$ (i.e., $\boldsymbol{s} = \boldsymbol{s}^*$).

Thus, for any solution $\boldsymbol{s}$ of the constraint equation in $(P_1)$, we have

$$\|\boldsymbol{s}\|_1 \geq \|\boldsymbol{s}^*\|_1 \tag{35}$$

the equality holds only when $\boldsymbol{s}$ is equal to $\boldsymbol{s}^*$. Thus, $\boldsymbol{s}_1 = \boldsymbol{s}^*$. The sufficiency is proved.

2. We now derive the second necessary and sufficient condition using the first one.

Necessity: Note that there are totally $2^l - 1$ optimization problems represented by (4) corresponding to $2^l - 1$ different subsets in $F$. Let $\boldsymbol{\delta}_I \ \left(= [\delta_{I1}, \ldots, \delta_{Im}]^T\right)$ and $\bar{\boldsymbol{\delta}}$ denote the optimal solutions of (4) and (5), respectively, $I_0$ denote the index set on which the entries of $\boldsymbol{s}^*$ and $\bar{\boldsymbol{\delta}}$ have the same sign. Obviously, $I_0 \in F$. It is not difficult to find that $\bar{\boldsymbol{\delta}} = \boldsymbol{\delta}_{I_0}$. Thus, if $\boldsymbol{s}_1 = \boldsymbol{s}^*$, according to the first necessary and sufficient condition in this theorem, we have

$$\sum_{k=1}^{m} [\mathrm{sign}(s_k^*)\bar{\delta}_k]_+ = \sum_{k \in I_0} |\bar{\delta}_k|$$

which is the optimal value (4) with the index set $I$ being $I_0$, is less than $\frac{1}{2}$.

The necessity is obtained.

Sufficiency: Note that $\forall I \in F, \boldsymbol{\delta}_I$ is a feasible solution of (5), and

$$\sum_{k=1}^{m} [\mathrm{sign}(s_k^*)\delta_{Ik}]_+ = \sum_{k \in I} |\delta_{Ik}|.$$

Thus, if

$$\sum_{k=1}^{m} [\mathrm{sign}(s_k^*)\bar{\delta}_k]_+ < \frac{1}{2}$$

then

$$\sum_{k=1}^{m} [\mathrm{sign}(s_k^*)\delta_{Ik}]_+ < \frac{1}{2}.$$

That is, all optimal values of the $2^l - 1$ optimization problems represented by (4) are less than $\frac{1}{2}$. From the the first necessary and sufficient condition, we have $\boldsymbol{s}_1 = \boldsymbol{s}^*$.

The sufficiency is obtained. □

## Appendix II
## Equivalence Between (4) and Linear Programming Problem

The optimization problem (4) is not a standard linear programming problem. However, it can be transformed into a standard linear programming problem as follows.

Let $G_0$ denote the index set $\{1, 2, \ldots, m\}$. For $k \in G_0 \setminus G$, define $\delta_k = u_k - v_k$, where $u_k, v_k \geq 0$. Let $\boldsymbol{A}_G$ denote the submatrix composed of all the columns of $\boldsymbol{A}$ with their column indices being in $G$, $\boldsymbol{A}_{G^c} = \boldsymbol{A} \backslash \boldsymbol{A}_G$, and let $\boldsymbol{\delta}_G$ denote the column vector composed of all entries of $\boldsymbol{\delta}$ with their indices being in $G$. It is not difficult to prove that the optimization problem (4) is equivalent to a standard linear programming problem

$$\begin{aligned}
&\min \left[ -\sum_{k \in G \backslash I} sign(s_k^*)\delta_k + \sum_{k \in G_0 \backslash G} (u_k + v_k) \right], \quad \text{s. t.} \\
&[\boldsymbol{A}_G, \boldsymbol{A}_{G^c}, -\boldsymbol{A}_{G^c}] \left[\boldsymbol{\delta}_G^T, \boldsymbol{u}^T, \boldsymbol{v}^T\right]^T = \boldsymbol{0} \\
&\sum_{k \in I} \mathrm{sign}(s_k^*)\delta_k - \sum_{k \in G \backslash I} \mathrm{sign}(s_k^*)\delta_k + \sum_{k \in G_0 \backslash G} (u_k + v_k) = 1, \\
&\mathrm{sign}(s_k^*)\delta_k > 0, \quad \text{for } k \in I \\
&\mathrm{sign}(s_k^*)\delta_k \leq 0, \quad \text{for } k \in G \setminus I \\
&\boldsymbol{u}, \boldsymbol{v} \geq 0.
\end{aligned} \tag{36}$$

Equation (36) can be rewritten as

$$\begin{aligned}
&\min \left[ -\sum_{k \in G \backslash I} \mathrm{sign}(s_k^*)\delta_k + \sum_{k \in G_0 \backslash G} (u_k + v_k) \right], \quad \text{s. t.} \\
&\begin{bmatrix} \boldsymbol{A}_G & \boldsymbol{A}_{G^c} & -\boldsymbol{A}_{G^c} \\ \boldsymbol{b}_1 & \boldsymbol{b}_2 & \boldsymbol{b}_2 \end{bmatrix} \left[\boldsymbol{\delta}_G^T, \boldsymbol{u}^T, \boldsymbol{v}^T\right]^T = \left[\boldsymbol{0}^T, 1\right]^T \\
&\mathrm{sign}(s_k^*)\delta_k > 0, \quad \text{for } k \in I \\
&\mathrm{sign}(s_k^*)\delta_k \leq 0, \quad \text{for } k \in G \setminus I \\
&\boldsymbol{u}, \boldsymbol{v} \geq 0
\end{aligned} \tag{37}$$

where $\boldsymbol{b}_1$ is a $\|\boldsymbol{s}^*\|_0$-dimensional row vector defined as follows: $b_{1k} = \mathrm{sign}(s_k^*)$ if $k \in I$, $b_{1k} = -\mathrm{sign}(s_k^*)$ if $k \in G \setminus I$. $\boldsymbol{b}_2$ is a $m - \|\boldsymbol{s}^*\|_0$-dimensional row vector of which each entry is one.

Since the sign function $\mathrm{sign}(s_k^*)$ is known, (37) is a standard linear programming problem.

## Appendix III
## The Proof of Lemma 1

*Proof* For any perturbation matrix sequence denoted as $\{\Delta \boldsymbol{A}_i\}$ of $\boldsymbol{A}$, suppose that $\lim_{i \to \infty} \|\Delta \boldsymbol{A}_i\| = 0$. Under the condition that the optimal solution $\bar{\boldsymbol{s}}$ of (8) has $n$ nonzero entries, to prove Lemma 1, it suffices to prove that (9) is feasible for all sufficiently small perturbation matrices $\Delta \boldsymbol{A}_i$, and

$$\lim_{i \to \infty} \|\bar{\boldsymbol{s}} - \bar{\boldsymbol{s}}_i\| = 0, \lim_{i \to \infty} |f(\bar{\boldsymbol{s}}) - f(\bar{\boldsymbol{s}}_i)| = 0 \tag{38}$$

where $\bar{\boldsymbol{s}}_i$ denotes the optimal solution of (9) under the perturbation matrix $\Delta \boldsymbol{A}_i$.

Suppose that the indices of nonzero entries of $\bar{\boldsymbol{s}}$ are $(i_1, \ldots, i_n)$, the matrix $\tilde{\boldsymbol{A}}$ is composed of $n$ columns of $\boldsymbol{A}$ with indices $(i_1, \ldots, i_n)$, the $n$ dimensional column vector $\tilde{\boldsymbol{s}}$ is composed of all nonzero entries of $\bar{\boldsymbol{s}}$. Thus, we have

$$\boldsymbol{x} = \tilde{\boldsymbol{A}}\tilde{\boldsymbol{s}}, \quad \tilde{\boldsymbol{s}} > \boldsymbol{0}. \tag{39}$$

Consider the equations

$$\boldsymbol{x} = (\tilde{\boldsymbol{A}} + \Delta\tilde{\boldsymbol{A}}_i)\tilde{\boldsymbol{s}}_i \tag{40}$$

where the perturbation matrix $\Delta\tilde{\boldsymbol{A}}_i$ is composed of the $n$ columns of $\Delta \boldsymbol{A}_i$ with indices $(i_1, \ldots, i_n)$. Noting that $\Delta\tilde{\boldsymbol{A}}_i$ is random, and $(\tilde{\boldsymbol{A}} + \Delta\tilde{\boldsymbol{A}}_i)$ is inverse with probability one, we have

$$\tilde{\boldsymbol{s}}_i = (\tilde{\boldsymbol{A}} + \Delta\tilde{\boldsymbol{A}}_i)^{-1}\boldsymbol{x}. \tag{41}$$

Since $\lim_{i\to\infty} \Delta\tilde{\boldsymbol{A}}_i = 0$, we have $\lim_{i\to\infty}(\tilde{\boldsymbol{A}} + \Delta\tilde{\boldsymbol{A}}_i)^{-1} = (\tilde{\boldsymbol{A}})^{-1}$ and $\lim_{i\to\infty} \tilde{\boldsymbol{s}}_i = \tilde{\boldsymbol{s}}$.

Define an $m$-dimensional column $\hat{\boldsymbol{s}}_i$ as follows: $\hat{s}_{ki} = \tilde{s}_{ji}$ if $k = i_j, j = 1, \ldots, n$; otherwise $\hat{s}_{ki} = 0$. We can conclude that if $i$ is sufficiently large, then $\tilde{\boldsymbol{s}}_i > \boldsymbol{0}$, and $\hat{\boldsymbol{s}}_i$ is a feasible solution of linear programming problem (9).

Note that for sufficiently large $i$, linear programming problem (9) is feasible under the perturbation matrix $\Delta\boldsymbol{A}_i$. Without loss of generality, we suppose that linear programming problem (9) is feasible under each perturbation matrix $\Delta\boldsymbol{A}_i$. For the above feasible solution sequence $\{\hat{\boldsymbol{s}}_i\}$, we have $\lim_{i\to\infty} \hat{\boldsymbol{s}}_i = \bar{\boldsymbol{s}}$. Since $f(\boldsymbol{s})$ is a linear function with respect to $\boldsymbol{s}$, $\lim_{i\to\infty} f(\hat{\boldsymbol{s}}_i) = f(\bar{\boldsymbol{s}})$.

We now prove that $\lim_{i\to\infty} \bar{\boldsymbol{s}}_i = \bar{\boldsymbol{s}}$.

Suppose that $\tilde{\bar{\boldsymbol{s}}}_i \in R^n$ is a column which is composed by all nonzero entries of $\bar{\boldsymbol{s}}_i$ (if $||\bar{\boldsymbol{s}}_i||_0 < n$, then $\tilde{\bar{\boldsymbol{s}}}_i$ is is composed by all nonzero entries of $\bar{\boldsymbol{s}}_i$ and zeros).

Similarly as in (41), nonnegative vector $\tilde{\bar{\boldsymbol{s}}}_i$ can be represented by $(\hat{\boldsymbol{A}}_i + \Delta\hat{\boldsymbol{A}}_i)^{-1}\boldsymbol{x}$, where $\hat{\boldsymbol{A}}_i$ and $\Delta\hat{\boldsymbol{A}}_i$ are square submatrices of $\boldsymbol{A}$ and $\Delta\boldsymbol{A}_i$, respectively.

First, we prove that the sequence $\{(\hat{\boldsymbol{A}}_i + \Delta\hat{\boldsymbol{A}}_i)^{-1}\boldsymbol{x}\}$ is bounded. Otherwise, there is a subsequence denoted as $\{(\hat{\boldsymbol{A}}_{i_j} + \Delta\hat{\boldsymbol{A}}_{i_j})^{-1}\boldsymbol{x}\}$, such that $\lim_{j\to\infty} ||(\hat{\boldsymbol{A}}_{i_j} + \Delta\hat{\boldsymbol{A}}_{i_j})^{-1}\boldsymbol{x}|| = \infty$. Since the number of $n \times n$-dimensional submatrices of $\boldsymbol{A}$ is $C_m^n$, the matrix sequence $\{\hat{\boldsymbol{A}}_{i_j}\}$ has a subsequence denoted as itself such that $\hat{\boldsymbol{A}}_{i_j} = \hat{\boldsymbol{A}}_0$ for all $j = 1, \ldots$, where $\hat{\boldsymbol{A}}_0$ is a $n \times n$-dimensional submatrix of $\boldsymbol{A}$. Since $\lim_{j\to\infty} ||\Delta\hat{\boldsymbol{A}}_{i_j}|| = 0$, we have

$$\lim_{j\to\infty} ||(\hat{\boldsymbol{A}}_{i_j} + \Delta\hat{\boldsymbol{A}}_{i_j})^{-1}\boldsymbol{x}|| = ||(\hat{\boldsymbol{A}}_0)^{-1}\boldsymbol{x}||.$$

A contradiction has happened. Hence, $\{(\hat{\boldsymbol{A}}_i + \Delta\hat{\boldsymbol{A}}_i)^{-1}\boldsymbol{x}\}$ is bounded.

Suppose that $\{(\hat{\boldsymbol{A}}_{l_k} + \Delta\hat{\boldsymbol{A}}_{l_k})^{-1}\boldsymbol{x}\}$ is a convergent subsequence of $\{(\hat{\boldsymbol{A}}_i + \Delta\hat{\boldsymbol{A}}_i)^{-1}\boldsymbol{x}\}$, and denote $\lim_{k\to\infty}(\hat{\boldsymbol{A}}_{l_k} + \Delta\hat{\boldsymbol{A}}_{l_k})^{-1}\boldsymbol{x} = \tilde{\bar{\boldsymbol{s}}}_0$. By adding zeros into $\tilde{\bar{\boldsymbol{s}}}_0$, we can obtain a column vector denoted as $\bar{\boldsymbol{s}}_0 \in R^m$. It is not difficult to check that $\bar{\boldsymbol{s}}_0$ is a feasible solution of (8), and $\lim_{k\to\infty} f(\bar{\boldsymbol{s}}_{l_k}) = f(\bar{\boldsymbol{s}}_0)$.

Moreover, since $\hat{\boldsymbol{s}}_{l_k}$ is a feasible solution of (9), thus,

$$f(\hat{\boldsymbol{s}}_{l_k}) \geq f(\bar{\boldsymbol{s}}_{l_k}) \quad \text{and} \quad \lim_{k\to\infty} f(\hat{\boldsymbol{s}}_{l_k}) \geq \lim_{k\to\infty} f(\bar{\boldsymbol{s}}_{l_k}).$$

That is, $f(\bar{\boldsymbol{s}}) \geq f(\bar{\boldsymbol{s}}_0)$. Since $\bar{\boldsymbol{s}}$ is the optimal solution of (8), which is unique with probability one, we have $\bar{\boldsymbol{s}}_0 = \bar{\boldsymbol{s}}$. From the preceding analysis, we find that any convergent subsequence of $\bar{\boldsymbol{s}}_i$ has the limit of $\bar{\boldsymbol{s}}$, thus, $\lim_{i\to\infty} \bar{\boldsymbol{s}}_i = \bar{\boldsymbol{s}}$. Furthermore, we have $\lim_{i\to\infty} f(\bar{\boldsymbol{s}}_i) = f(\bar{\boldsymbol{s}})$.

The lemma is proved. □

## Appendix IV
## The Proof of Theorem 3

*Proof* According to (7), the probability $P(\boldsymbol{s}^* = \boldsymbol{s}_1; ||\boldsymbol{s}^*||_0 = l, \boldsymbol{A})$ can be determined by the recoverability of the $N$ sign vectors $\boldsymbol{t}(1), \ldots, \boldsymbol{t}(N)$, which have exactly $l$ nonzero entries, where $N = 2^l C_m^l$. We consider the case in which the $N$ sign vectors are source vectors. For a given matrix $\boldsymbol{A}_0$, suppose that all the 1-norm solutions of the following $2N$ optimization problems are unique:

$$\min \sum_{i=1}^{m} |s_i|, \quad \text{s. t. } \boldsymbol{A}_0\boldsymbol{s} = \boldsymbol{A}_0\boldsymbol{t}(q) \tag{42}$$

$$\min \sum_{i=1}^{m} |s_i|, \quad \text{s. t. } (\boldsymbol{A}_0 + \Delta\boldsymbol{A})\boldsymbol{s} = (\boldsymbol{A}_0 + \Delta\boldsymbol{A})\boldsymbol{t}(q), \tag{43}$$

where $q = 1, \ldots, N$.

Note that the set of $\boldsymbol{A}_0$, under which at least one of the above optimization problems has more than one solution, has measure zero.

Suppose that for the given $\boldsymbol{A}_0$, there are $N_1$ sign vectors which can be recovered, that is, $P(\boldsymbol{s}^* = \boldsymbol{s}_1; ||\boldsymbol{s}^*||_0 = l, \boldsymbol{A}_0) = \frac{N_1}{N}$. We consider the optimization problem

$$\begin{aligned} &\max \sum_{k\in I} |\delta_k|, \quad \text{s. t.} \\ &\boldsymbol{A}_0\boldsymbol{\delta} = 0, \quad ||\boldsymbol{\delta}||_1 = 1 \\ &\delta_k t_k(q) > 0, \quad \text{for } k \in I, \\ &\delta_k t_k(q) \leq 0, \quad \text{for } k \in G \setminus I. \end{aligned} \tag{44}$$

Suppose that $\boldsymbol{t}(q)$ can be recovered by solving the linear programming problem $(P_1)$. Then $\forall I \in F$, when the optimization problem (44) is solvable, its optimal value is less than $\frac{1}{2}$.

Now we consider a corresponding optimization problem

$$\begin{aligned} &\max \sum_{k\in I} |\delta_k|, \quad \text{s. t.} \\ &(\boldsymbol{A}_0 + \Delta\boldsymbol{A})\boldsymbol{\delta} = 0, \quad ||\boldsymbol{\delta}||_1 = 1 \\ &\delta_k t_k(q) > 0, \quad \text{for } k \in I \\ &\delta_k t_k(q) \leq 0, \quad \text{for } k \in G \setminus I. \end{aligned} \tag{45}$$

Note that the preceding two optimization problems are equivalent to standard linear programming problems which are similar to (37) in Appendix II. For the linear programming problem (37) with $n+1$ equality constraints, it is not difficult to prove that any $(n+1) \times (n+1)$-dimensional submatrix of its constraint coefficient matrix is of full rank. Furthermore, if (37) is solvable, then its optimal solution has $n+1$ nonzero entries. Thus, if optimization problem (44) is solvable, then the optimal solution of its equivalent linear programming problem has $n+1$ nonzero entries. From Lemma 1, if the optimization value of (44) is less than $\frac{1}{2}$, then the optimization value of (45) is also less than $\frac{1}{2}$ when $||\Delta\boldsymbol{A}||$ is sufficiently small, and *vice versa*. In other words, if $||\Delta\boldsymbol{A}||$ is sufficiently small, the sign vectors, which can be recovered by solving the problem (42), are the same as those which can be recovered by solving the problem (43). Thus,

$$\begin{aligned} &P\{\boldsymbol{s}^* = \boldsymbol{s}_1; ||\boldsymbol{s}^*||_0 = l, \boldsymbol{A}_0\} \\ &\qquad = P\{\boldsymbol{s}^* = \boldsymbol{s}_1; ||\boldsymbol{s}^*||_0 = l, \boldsymbol{A}_0 + \Delta\boldsymbol{A}\} = \frac{N_1}{N} \end{aligned}$$

if $||\Delta\boldsymbol{A}||$ is sufficiently small. The theorem is proved. □

## Appendix V
## The Proof of Theorem 4

For each sign vector $\boldsymbol{t}$, let us define the subset of $S = \{\boldsymbol{A}\}$

$$S(\boldsymbol{t}) = \left\{\boldsymbol{A} | M(\boldsymbol{A}, \boldsymbol{t}) < \frac{1}{2}\right\} \tag{46}$$

which is the set of matrices $\boldsymbol{A}$ for which the source $\boldsymbol{t}$ is recoverable by the $L_1$-solution. The recoverability probability of a source $\boldsymbol{s}^*$ whose sign vector is $\boldsymbol{t}$ is written as

$$P(\boldsymbol{t}) = \text{Prob}\{\boldsymbol{A} \in S(\boldsymbol{t})\} = \text{Prob}\left\{\boldsymbol{A} | M(\boldsymbol{A}, t) < \frac{1}{2}\right\}. \tag{47}$$

There are $N$ such subsets $S(\boldsymbol{t}(1)), \ldots, S(\boldsymbol{t}(N))$.

Let us study invariant properties of $M(\boldsymbol{A}, \boldsymbol{t})$. Let $\prod$ be the group consisting of permutations of $m$ elements $(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_m)$ or $(t_1, \ldots, t_m)$ and change of some of their signs. Its element $T$ act on $\boldsymbol{A}$ or $\boldsymbol{t}$ as

$$T\boldsymbol{A} = \left(s_{i(1)}\boldsymbol{a}_{i(1)}, \ldots, s_{i(m)}\boldsymbol{a}_{i(m)}\right) \tag{48}$$

$$T(\boldsymbol{t}) = \left(s_{i(1)}t_{i(1)}, \ldots, s_{i(m)}t_{i(m)}\right) \tag{49}$$

where $\{i(1), \ldots, i(m)\}$ is a permutation of $\{1, \ldots, m\}$ by $T$ and $(s_1, \ldots, s_m)$ is the change of signs, $s_i = 1$ or $-1$, and the element with $s_i = -1$ changes sign.

It is obvious to see that

$$(T\boldsymbol{A})(T\boldsymbol{\delta}) = \boldsymbol{A\delta} \tag{50}$$

$$\|\boldsymbol{\delta}\|_1 = \|T\boldsymbol{\delta}\|_1 \tag{51}$$

hold. The condition $\delta_k s_k^* > 0,\ k \in F$ or $\delta_k s_k^* \leq 0$ for $k \in G \backslash I$ is also equivariant under $T$.

Hence, $M(\boldsymbol{A}, \boldsymbol{t})$ is invariant under $\prod$, that is,

$$M(\boldsymbol{A}, \boldsymbol{t}) = M(T\boldsymbol{A}, T\boldsymbol{t}). \tag{52}$$

Furthermore, since each element $a_{ij}$ is independently and symmetrically distributed, i.e., $P(a_{ij}) = P(-a_{ij})$, thus, probability density $P(\boldsymbol{A})$ is invariant under $\prod$, that is.

$$P(\boldsymbol{A}) = P(T\boldsymbol{A}) \tag{53}$$

for any $T \in \prod$.

Given sign vector $\boldsymbol{t}$, we introduce a related element $T(\boldsymbol{t}) \in \prod$.

Let $t_{i_1}, \ldots, t_{i_l}$ be the nonzero elements of $\boldsymbol{t}$. Then $T(\boldsymbol{t})$ locates $l$ elements $(i_1, \ldots, i_l)$ to the first $l$ positions, and changes the sign of $i_j$ elements when $t_{i_j} = -1$. It locates the other $m - l$ elements randomly in the positions following the first $l$ elements. This implies that

$$T(\boldsymbol{t})\boldsymbol{t} = \boldsymbol{t}(1) = (1, 1, \ldots, 1, 0, \ldots, 0) \tag{54}$$

for any $\boldsymbol{t}$. Let $I(\boldsymbol{A}, \boldsymbol{t})$ be the indicator function such that $(\boldsymbol{A}, \boldsymbol{t})$ is recoverable, that is, $I(\boldsymbol{A}, \boldsymbol{t}) = 1$ when $M(\boldsymbol{A}, \boldsymbol{t}) < 1/2$ and 0 otherwise.

Since the optimization problem (4) is invariant under permutations of columns of $\boldsymbol{A}$ and sign changes, the recoverability of $(\boldsymbol{A}, \boldsymbol{t})$ is the same as that of $(T(\boldsymbol{t})\boldsymbol{A}, \boldsymbol{t}(1))$, and hence,

$$I(\boldsymbol{A}, \boldsymbol{t}) = I(T(\boldsymbol{t})\boldsymbol{A}, \boldsymbol{t}(1)) \tag{55}$$

for all $\boldsymbol{t}$.

This leads us to the conclusion of the theorem, that is, for any $\boldsymbol{t}$, $P(\boldsymbol{t})$ is the same, and

$$P(l, \boldsymbol{A}) = P(\boldsymbol{t}) \tag{56}$$

that is, the recoverability probability does not depend on $\boldsymbol{t}$.

Hence, Theorem 4 is proved. □

## Appendix VI
## Sketch of Proof of Proposition 1

*Proof* First, we denote the set of feasible solutions of (5) as $D\ (= \{\boldsymbol{\delta} |\ \boldsymbol{A\delta} = \boldsymbol{0},\ \|\boldsymbol{\delta}\|_1 = 1\})$. For a sign vector $\boldsymbol{t}(q)$, let $\bar{\boldsymbol{\delta}}_q$ denote the optimal solution of (5), where the source vector $\boldsymbol{s}^* = \boldsymbol{t}(q)$. The index set of nonzero entries of $\boldsymbol{t}(q)$ is denoted as $G_q$.

We now consider the case in which $l$, the number of nonzero entries of each sign vector, is small (much less than $n$ or close to 1). Note that small $l$ implies that the $N$ sign vectors are very sparse.

For the optimal solution $\bar{\boldsymbol{\delta}}_q$, the objective function value of (5) becomes $\sum_{k \in G_q} [t_{kq}\bar{\delta}_{kq}]_+$, which is a summation of at most $l$ items. Since $l$ is small, and $\sum_{k \in G_q} [t_{kq}\bar{\delta}_{kq}]_+$ is the maximum of $\sum_{k \in G_q} [t_{kq}\delta_{kq}]_+$ for $\boldsymbol{\delta} \in D$, thus generally we have $[t_{kq}\bar{\boldsymbol{\delta}}_{kq}]_+ > 0$, i.e., $t_{kq}\bar{\boldsymbol{\delta}}_{kq} > 0$ for $k \in G_q$.

Thus, for the source vector $\boldsymbol{t}(q)$, optimization problem (5) becomes

$$\max \sum_{k \in G_q} t_{kq}\delta_k, \quad \text{s. t.}$$
$$\boldsymbol{A\delta} = 0, \quad \|\boldsymbol{\delta}\|_1 = 1. \tag{57}$$

Note: (57) was proposed in [27], and used as a conjecture in [28].

Since sign vectors $\boldsymbol{t}(q)$, $q = 1, \ldots, N$ are different, the objective functions $\sum_{k \in G_q} t_{kq}\delta_k$ of (57) are different. Obviously, the optimal value of (57) is equal to $\max\limits_{\boldsymbol{\delta} \in D} \sum_{k \in G_q} t_{kq}\delta_k$. Since the set of solutions of $\boldsymbol{A\delta} = 0$ is an $m - n$-dimensional linear space, of which each solution can be in the feasible solution set $D$ through normalization, $D$ is a compact set in the $m - n$-dimensional space.

For sign vector $\boldsymbol{t}(q)$, define a subset of $D$ as

$$D_q = \{\boldsymbol{\delta} | \boldsymbol{\delta} \in D, t_{kq}\delta_k > 0 \text{for}, k \in G_q\}.$$

It is not difficult to find that $D_q$ is also a compact set in $m - n$-dimensional linear space.

For two different sign vectors $\boldsymbol{t}(q_1)$ and $\boldsymbol{t}(q_2)$, we consider three cases.

1. $G_{q_1} \bigcap G_{q_2} \neq \phi$.

If $\boldsymbol{t}(q_1)$ and $\boldsymbol{t}(q_2)$ have different elements with their indices in $G_{q_1} \bigcap G_{q_2}$, then $D_{q_1} \bigcap D_{q_2} = \phi$. Furthermore, we have

$$\max_{\boldsymbol{\delta} \in D} \sum_{k \in G_q} t_{kq}\delta_k = \max_{\boldsymbol{\delta} \in D_q} \sum_{k \in G_q} t_{kq}\delta_k.$$

Thus the two events

$$\left(\max_{\boldsymbol{\delta} \in D} \sum_{k \in G_{q_1}} t_{kq_1}\delta_k < \frac{1}{2}\right) \quad \text{and} \quad \left(\max_{\boldsymbol{\delta} \in D} \sum_{k \in G_{q_2}} t_{kq_2}\delta_k < \frac{1}{2}\right)$$

are close to being uncorrelated. Noting that

$$\left(\max_{\boldsymbol{\delta}\in D}\sum_{k\in G_q} t_{kq}\delta_k < \frac{1}{2}\right)$$

is equivalent to $\left(M(\boldsymbol{A},\boldsymbol{t}(q)) < \frac{1}{2}\right)$ here, $\left(M(\boldsymbol{A},\boldsymbol{t}(q_1)) < \frac{1}{2}\right)$ and $\left(M(\boldsymbol{A},\boldsymbol{t}(q_2)) < \frac{1}{2}\right)$ are close to being uncorrelated.

On index set $G_{q_1}\bigcap G_{q_2}$, if $\boldsymbol{t}(q_1)$ and $\boldsymbol{t}(q_2)$ share the same elements, then $D_{q_1}\bigcap D_{q_2}$ may not be null set. Considering the two objective functions $\sum_{k\in G_{q_1}} t_{kq_1}\delta_k$ and $\sum_{k\in G_{q_2}} t_{kq_2}\delta_k$ of (57) corresponding to $\boldsymbol{t}(q_1)$ and $\boldsymbol{t}(q_2)$, respectively, are different (but share several common items), the two events

$$\left(\max_{\boldsymbol{\delta}\in D}\sum_{k\in G_{q_1}} t_{kq_1}\delta_k < \frac{1}{2}\right) \quad\text{and}\quad \left(\max_{\boldsymbol{\delta}\in D}\sum_{k\in G_{q_2}} t_{kq_2}\delta_k < \frac{1}{2}\right)$$

may be close to being uncorrelated. Even it is not true, compared with the total number of $N(N-1)$ of sign vector pairs, the number of such pairs $(\boldsymbol{t}(q_1),\boldsymbol{t}(q_2))$ is small.

2. $G_{q_1}\bigcap G_{q_2}=\phi$.

In this case, the two objective functions $\sum_{k\in G_{q_1}} t_{kq_1}\delta_k$ and $\sum_{k\in G_{q_2}} t_{kq_2}\delta_k$ of (57) corresponding to $\boldsymbol{t}(q_1)$ and $\boldsymbol{t}(q_2)$, respectively, do not share any common variable. Since $D$ is a compact set in the $m-n$-dimensional space, the two events

$$\left(\max_{\boldsymbol{\delta}\in D}\sum_{k\in G_{q_1}} t_{kq_1}\delta_k < \frac{1}{2}\right) \quad\text{and}\quad \left(\max_{\boldsymbol{\delta}\in D}\sum_{k\in G_{q_2}} t_{kq_2}\delta_k < \frac{1}{2}\right)$$

are close to being uncorrelated.

Combining the analysis in above two cases, we can say most of event pairs

$$\left(\max_{\boldsymbol{\delta}\in D}\sum_{k\in G_{q_1}} t_{kq_1}\delta_k < \frac{1}{2}\right) \quad\text{and}\quad \left(\max_{\boldsymbol{\delta}\in D}\sum_{k\in G_{q_2}} t_{kq_2}\delta_k < \frac{1}{2}\right)$$

are close to being uncorrelated. We have obtained the conclusion in Proposition 1 when $l$ is small.

As $l$ increases, for some mixing matrix $\boldsymbol{A}$ and sign vectors (source vectors), (5) may not be equivalent to (57). Thus, for different sign vectors $\boldsymbol{t}(q_1)$ and $\boldsymbol{t}(q_2)$, and some mixing matrix $\boldsymbol{A}$, their corresponding optimization problems represented by (5) may have the same optimal solution. Consequently, the correlation of event pair of

$$\left(\max_{\boldsymbol{\delta}\in D}\sum_{k\in G_{q_1}} t_{kq_1}\delta_k < \frac{1}{2}\right) \quad\text{and}\quad \left(\max_{\boldsymbol{\delta}\in D}\sum_{k\in G_{q_2}} t_{kq_2}\delta_k < \frac{1}{2}\right)$$

may increase. It seems that the larger $l$, the higher correlation. However, if $l$ is close or equal to $n$, generally, the probability that a sign vector is recovered is close to zero. Thus, for any 2-sign vectors $\boldsymbol{t}(q_1),\boldsymbol{t}(q_2)$, we have

$$\begin{aligned} &P\left(M\left(\boldsymbol{A},\boldsymbol{t}(q_1)\right)<\frac{1}{2}, M\left(\boldsymbol{A},\boldsymbol{t}(q_2)\right)<\frac{1}{2}\right)\\ &\approx P\left(M\left(\boldsymbol{A},\boldsymbol{t}(q_1)\right)P(M(\boldsymbol{A},\boldsymbol{t}(q_2))<\frac{1}{2}\right)\\ &\approx 0. \end{aligned} \tag{58}$$

Hence, the any two events $\left(M(\boldsymbol{A},\boldsymbol{t}(q_1)) < \frac{1}{2}\right)$ and $\left(M(\boldsymbol{A},\boldsymbol{t}(q_2)) < \frac{1}{2}\right)$ are still close to being uncorrelated.

Finally, if $l>n$, no sign vector can be recovered. All probabilities in (58) are 0, and the two "≈" become "=," thus any two events are independent. □

## APPENDIX VII
## PROOF OF LEMMA 2

We define

$$B[k]=\frac{1}{k}\sum_{i=1}^{k}x[i]. \tag{59}$$

Let $e$ be the mean of $x[i]$. What we want to show is

$$\lim_{k\to\infty} E[(B[k]-e)^2]=0 \tag{60}$$

implying that $B[k]$ converges to the mean value in the sense of mean-square error (that is, in the weak sense).

Note that $E[(B[k]-m)^2]$ in (60) is the variance of $B[k]$. We now give a sufficient condition for (60) (Obviously, that $x[i]$ are independent is a sufficient condition). To this end, we calculate the variance of $B[k]$. Here we assume, without loss of generality, $e=0$. (If it is not 0, we just consider $x[i]-e$ as new random variables.)

We have

$$Var[B[k]]=(1/k^2)\left[\sum E[x[i]^2]+\sum E[x[i]x[j]]\right]. \tag{61}$$

There are $k$ terms in the first sum, so it converges to 0 if divided by $k^2$. There are $k(k-1)$ in the second, which are covariances of $x[i]$ and $x[j]$. So if this covariances converges to 0 for "most" $i$ and $j$, or more precisely, converges to 0 except for those special exceptional terms whose number is of order $k$, the Lemma 2 is proved. □.

## REFERENCES


[1] S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Scientific Computing*, 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] D. L. Donoho and D. Neighborly, Polytopes and Sparse Solution of Underdetermined Linear Equations 2005 [Online]. Available: http://www-stat.stanford.edu/donoho/Reports/2005/NPaSSULE-01-28-05.pdf

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

[5] 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.

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

[7] 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.

[8] ——, "Greed is good: Algorithmic results for sparse approximation," *IEEE Trans. Inf. Theory*, vol. 50, no. 10, pp. 2231–2242, 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. Acoustics, Speech and Signal Processing*, Philadelphia, PA, Mar. 2005, vol. V, pp. 717–720.

[10] J. A. Tropp, A. C. Gilbert, S. Muthukrishnan, and M. J. Strauss, "Improved sparse approximation over quasiincoherent dictionaries," in *Proc. 2003 IEEE Int. Confe. Image Processing*, Barcelona, Spain, Sep. 2003, pp. 37–40.

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

[12] J. Eriksson and V. Koivunen, "Identifiability and separability of linear ICA models revisited," in *Proc. 4th Int. Symp. Independent Component Analysis and Blind Signal Separation*, Nara, Japan, 2003, pp. 23–27.

[13] M. Davies, “Identifiability issues in noisy ICA,” *IEEE Signal Process. Lett.*, vol. 11, no. 5, pp. 470–473, May 2004.

[14] J. -F. Cardoso, “Blind signal separation: Statistical principles,” *Proc. IEEE*, vol. 86, no. 10, pp. 2009–2025, Oct. 1998.

[15] Y. Li and J. Wang, “Sequential blind extraction of linearly mixed sources,” *IEEE Trans. Signal Process.*, vol. 50, no. 5, pp. 997–1007, May 2002.

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

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

[18] M. Zibulevsky, B. A. Pearlmutter, P. Bofill, and P. Kisilev, , S. J. Roberts and R. M. Everson, Eds., “Blind source separation by sparse decomposition in a signal dictionary,” in *Independent Components Analysis: Principles and Practice*. Cambridge, U.K.: Cambridge Univ. Press, 2000.

[19] 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.

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

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

[22] J.-J. Fuchs, “Recovery of exact sparse representations in the presence of noise,” in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, Montreal, QC, Canada, 2004, pp. II–533–536.

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

[24] D. M. Malioutov, M. Cetin, and A. S. Willsky, “Optimal sparse representations in general overcomplete bases,” in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, Montreal, QC, Canada, May 2004, vol. 2, pp. II-793–796.

[25] D. L. Donoho, For Most Large Underdetermined Systems of Linear Equations, the Minimal L1-Norm Nearsolution Approximates the Sparsest Near-Solution. (Report) [Online]. Available: http://www-stat.stanford.edu/donoho/reports.html

[26] ——, For Most Large Underdetermined Systems of Linear Equations, the Minimal L1-Norm Solution is Also the Sparsest Solution. (Report) [Online]. Available: http://www-stat.stanford.edu/donoho/reports.html

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

[28] M. Elad and M. Zibulevsky, A Probabilistic Study of the Average Performance of the Basis Pursuit [Online]. Available: http://www.cs.technion.ac.il/elad/publications/journals/