Underdetermined Blind Separation by Combining Sparsity and Independence of Sources

**PENG CHEN[1], DEZHONG PENG [1], (Member, IEEE), LIANGLI ZHEN[1], YIFAN LUO[1],
AND YONG XIANG [2], (Senior Member, IEEE)**

[1]College of Computer Science, Sichuan University, Chengdu 610065, China
[2]School of Information Technology, Deakin University, Burwood VIC 3125, Australia

Corresponding author: Yong Xiang (yxiang@deakin.edu.au)

ABSTRACT In this paper, we address underdetermined blind separation of N sources from their M instantaneous mixtures, where $N > M$, by combining the sparsity and independence of sources. First, we propose an effective scheme to search some sample segments with the local sparsity, which means that in these sample segments, only $Q(Q < M)$ sources are active. By grouping these sample segments into different sets such that each set has the same Q active sources, the original underdetermined BSS problem can be transformed into a series of locally overdetermined BSS problems. Thus, the blind channel identification task can be achieved by solving these overdetermined problems in each set by exploiting the independence of sources. In the second stage, we will achieve source recovery by exploiting a mild sparsity constraint, which is proven to be a sufficient and necessary condition to guarantee recovery of source signals. Compared with some sparsity-based UBSS approaches, this paper relaxes the sparsity restriction about sources to some extent by assuming that different source signals are mutually independent. At the same time, the proposed UBSS approach does not impose any richness constraint on sources. Theoretical analysis and simulation results illustrate the effectiveness of our approach.

INDEX TERMS Underdetermined blind source separation, sparsity, independence, source recovery, blind identification.

I. INTRODUCTION

Blind source separation (BSS) aims to separate source signals from their measurable mixtures without any *prior* knowledge about the mixing matrix and the source signals [1]–[15]. A challenging issue in BSS is that the number of sources exceeds the number of mixtures, which refers to the so-called underdetermined blind source separation (UBSS). This situation is often encountered in practical applications. For example, in a wireless sensor network, the number of sources is sometimes unknown to the receivers. As a result, the number of deployed receiving sensors could be less than the number of the sources. Other applications of UBSS include medical imaging, time series analysis and so on [10]. Therefore, UBSS is of particular importance in practice and has attracted considerable research attentions in recent years.

So far, various UBSS methods have been developed. Among these methods, a small number of them employs some special properties of the source signals and/or the mixing matrix. In [11] and [12], the differences between nonstationary source signals and stationary source signals are exploited to separate the former from the measured mixtures, whilst the latter are treated as noise signals and thus cannot be recovered. Differently, [14] uses a special property of the mixing matrix to develop some sequential underdetermined blind extraction techniques. Specifically, in order to estimate one single source signal from the mixtures, it is required that there must exist at least one column in the mixing matrix which is not a linear combination of all other columns [14]. This constraint is restrictive in practical applications since any column in a randomly generated underdetermined matrix can be expressed, with probability one, as a linear combination of all other columns. Some high-order cumulant-based blind identification approaches are developed in [4] and [5], which cannot recover source signals, and impose some restrictions on the maximum number of outputs.

On the other hand, a large number of UBSS methods perform source separation by utilizing the sparsity properties of sources in time domain [16]–[19] or other transformation domains, such as time-frequency domain [10], [20]–[26]. Sparse representation provides a powerful tool for solving the UBSS problem [15], [19], [33]. Underdetermined blind source separation is achieved in [16]–[18], [20] by assuming that at most one source signal possesses the dominant energy at each sample point, which is called as the W-disjoint orthogonal condition (or approximate W-disjoint orthogonality). For example, a so-called DUET algorithm is proposed in [20], [21] to estimate the underdetermined mixing matrix by exploiting the ratios of the time-frequency transforms of the observed mixture signals under the above W-disjoint orthogonal condition. In [29]–[31], the sparsity constraint about source signals is relaxed to some extent by allowing the energy distributions of sources to overlap. Specifically, it is not required in [29]–[31] that each sample point has at most one active source, and a weaker sparsity condition is presented, i.e., that there are some adjacent sample regions where only one source occurs. Based on this condition, the TIFROM algorithm is proposed to tackle the UBSS task in [29]–[31]. More recently, as an extension of the DUET algorithm and the TIFROM algorithm, an effective UBSS algorithm is developed by [15] to obtain the estimation for the mixing matrix under a milder sparsity condition, i.e., that there exist some isolated (or discrete) sample points on which only one single source is active. Clearly, it is more relax than the sparsity constraints required by the DUET algorithm and the TIFROM algorithm. It is notable that all of the above-mentioned sparsity constraints need some sample points having at most one active source signal, that is, there must exist some sample points where the energy distributions of source signals do not overlap. Such a sparsity restriction could be too difficult to satisfy in some practical applications, where all sample points have more than one active source. Interestingly, this restriction is successfully removed in [19] at the cost of an additional condition about the richness of the source signals. Specifically, [19] allows the distributions of source signals to overlap on all sample points, and does not need any sample points with only one active source. However, as an additional richness constraint, the sources are required to be sufficiently rich such that for any $N - M + 1$ sources from all N sources, there must exist at least M sample points on which these $N - M + 1$ sources are inactive [19], where N and M stand for the number of sources and the number of mixtures, respectively. It is notable that the richness condition is still difficult to satisfy in some practical applications, which will be further illustrated by an example presented in Section II.

In order to overcome the drawbacks of the existing works, this paper presents a group-based UBSS algorithm. Specifically, we group all the available mixture vectors into a number of sets such that each set has only $Q(Q < M)$ active sources. As a result, the UBSS problem is transformed into a series of locally non-underdetermined BSS problems. By solving these locally non-underdetermined BSS problems, we can achieve UBSS. The above idea is similar to that of [32], which combines the sparsity and independence of sources to solve the UBSS problem. Both our proposed algorithm and [32] need to group all the observed mixture vectors into different sets, by which the original UBSS task can be achieved by solving a series of non-determined BSS problems in each set. The main difference between our algorithm and [32] lies in that different methods are used in the above grouping procedure. In order to obtain such one group, a searching procedure regarding a particular cost function has to be executed based on a steep ascent method in [32]. Clearly, this procedure will suffer from the problem of local maxima. Although [32] proposed some amendments for this problem, this adds greatly the computational complexity. In contrast, our proposed algorithm only need an SVD operation and some simple comparisons for getting each group. On the other hand, our proposed method can estimate the mixing matrix, and then recover source signals, while [32] only can achieve the estimation for the mixing matrix. It is well-known that even though the mixing matrix is known, it is not a trivial task to recover source signals in the underdetermined system, where the mixing matrix is not invertible. It is notable that some blind source recovery algorithms have been developed in [10] and [26] by exploiting the known mixing matrix. For example, [26] exploits the estimated mixing matrix and subspace projection to identify the sources present in each sample point of time-frequency (TF) plane, and then estimates their corresponding time-frequency distribution values. Finally, source signals can be recovered via a TF synthesis algorithm [26]. More recently, [10] points out that [26] does not provide the sufficient conditions to guarantee the effectiveness of the UBSS algorithm, and relaxes the related sparsity restriction about source signals compared with [26]. However, [10] depends on the quadratic time-frequency distributions of signals, which makes the performance of the UBSS algorithm in [10] be negatively affected by the cross terms among different sources.

In this paper, we will propose a UBSS algorithm, which relaxes the sparsity assumption about source signals by exploiting the statistical independence among different source signals, and permits all sample points to possess more than one active source. Besides, unlike [19], the proposed algorithm does not need any additional constraint on the richness of the source signals. The advantages of the proposed approach make it applicable to a wider range of practical applications.

The rest of this paper is organized as follows. In Section II, the UBSS problem is formulated and the UBSS algorithm in [19] is reviewed. Section III presents the new UBSS algorithm based on some mild constraints on the sources and the mixing matrix. In Section IV, simulation results are provided to illustrate the effectiveness of the proposed approach. Finally, Section V concludes the paper.

Notations: The bold-faced upper case letters denote matrices (e.g., $\mathbf{X}$). The bold-faced lower case letters represent column vectors (e.g., $\mathbf{x}$) and x_i is the ith element of the

vector $\mathbf{x}$. The superscript $(\cdot)^T$ denotes the transpose operation. $\mathbf{I}_{M \times M}$ and $\mathbf{0}_{M \times 1}$ stand for the $M \times M$ unit matrix and the M-dimensional zero vector, respectively. pinv $(\mathbf{X})$, rank $[\mathbf{X}]$ and span $\{\mathbf{X}\}$ denote the pseudoinverse of the matrix $\mathbf{X}$, the rank of $\mathbf{X}$ and the subspace spanned by all the column vectors of $\mathbf{X}$, respectively.

II. PROBLEM FORMULATION AND REVIEW OF RELATED WORKS

Let us consider an instantaneous underdetermined mixing system with N sources and M outputs ($N > M$):

$$\mathbf{x}(k) = \mathbf{A}\mathbf{s}(k) + \mathbf{n}(k) \tag{1}$$

where $\mathbf{s}(k) = [s_1(k), s_2(k), \ldots, s_N(k)]^T$ is the N-dimensional source signal vector, $\mathbf{x}(k) = [x_1(k), x_2(k), \ldots, x_M(k)]^T$ is the M-dimensional mixture signal vector, $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_N]$ is the $M \times N$ mixing matrix, $\mathbf{n}(k)$ is the additive noise vector, and $k = 1, 2, \ldots, K$. In (1), k is an index given to each sample of signals. As a generalized index, k has different meanings in different domains. For example, k stands for one time instant in the time domain, while in the time-frequency (TF) domain, the index k denotes one sample point in TF plane. The mixing system (1) has been widely used in the study of UBSS [10]–[15], [17]–[26]. The column vector $\mathbf{a}_i$ in $\mathbf{A}$ is called the steering vector corresponding to the ith source signal $s_i(k)$, where $i = 1, 2, \ldots, N$ [10]. The objective of UBSS is to recover the source signals $\mathbf{s}(k)$ only from their observed mixture signals $\mathbf{x}(k)$. Since $N > M$, the mixing matrix $\mathbf{A}$ is not invertible, which makes UBSS a very challenging task.

To perform UBSS, a large number of approaches have been developed based on different sparsity assumptions about source signals. Most of these sparsity-based UBSS algorithms require the existence of some adjacent or discrete sample points, on which only one source is active. By searching these sample points, one can achieve the estimation for the mixing matrix, and then recover source signals. More interestingly, In [19], Georgiev *et al.* relax the sparsity constraint to develop an effective UBSS algorithm, which does not need any sample points with only one active source. In other words, it allows multiple source to be active on all sample points. Unfortunately, [19] requires an additional restriction about the richness of the source signals, that is,

A1) The source signals are sufficiently rich in the sense that for any $N - M + 1$ sources in all N sources, there must exist at least M sample points at which these $N - M + 1$ sources are inactive.

It is notable that the above assumption A1) is still strict and difficult to satisfy in many practical applications. To illustrate this point, we consider a simple UBSS example with $M = 3$ outputs and $N = 4$ source signals $s_1(k), s_2(k), s_3(k), s_4(k)$, which means that the mixing matrix $\mathbf{A}$ is of dimension 3×4. According to the assumption A1), these 4 source signals must satisfy the following $C_N^{N-M+1} = C_4^2 = 6$ conditions:

c1) There are at least $M = 3$ time instants at which $s_1(k)$ and $s_2(k)$ are inactive.

c2) There are at least $M = 3$ time instants at which $s_1(k)$ and $s_3(k)$ are inactive.

c3) There are at least $M = 3$ time instants at which $s_1(k)$ and $s_4(k)$ are inactive.

c4) There are at least $M = 3$ time instants at which $s_2(k)$ and $s_3(k)$ are inactive.

c5) There are at least $M = 3$ time instants at which $s_2(k)$ and $s_4(k)$ are inactive.

c6) There are at least $M = 3$ time instants at which $s_3(k)$ and $s_4(k)$ are inactive.
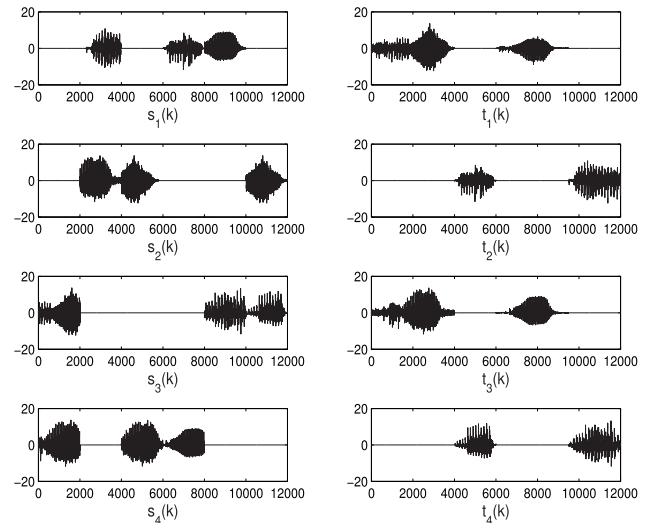

FIGURE 1. **Comparisons between two sets of source signals. Left column: source signal set $\{s_1(k), s_2(k), s_3(k), s_4(k)\}$ satisfying the assumption A1). Right column: source signal set $\{t_1(k), t_2(k), t_3(k), t_4(k)\}$ violating the assumption A1).**

Regarding the above six conditions, let us consider two sets of source signals $\{s_1(k), s_2(k), s_3(k), s_4(k)\}$ and $\{t_1(k), t_2(k), t_3(k), t_4(k)\}$, which are shown in the left and right columns in Fig. 1, respectively. It is easy to see that the source signals $\{s_1(k), s_2(k), s_3(k), s_4(k)\}$ satisfy all the above conditions c1)-c6) but the source signals $\{t_1(k), t_2(k), t_3(k), t_4(k)\}$ only meet the conditions c2) and c5). Specifically, for the source signals $\{t_1(k), t_2(k), t_3(k), t_4(k)\}$, the condition c5) is met during the time instants 1–4000, whilst the condition c2) is satisfied during the time instants 5000–6000. However, the conditions c1), c3), c4), and c6) cannot be satisfied by $\{t_1(k), t_2(k), t_3(k), t_4(k)\}$. As a result, $\{t_1(k), t_2(k), t_3(k), t_4(k)\}$ cannot be separated from their underdetermined mixtures by using the algorithm in [19]. From this example, we can see that the assumption A1) seriously limits the application scope of the UBSS algorithm in [19]. It is also worth mentioning that the assumption required by [16]–[18], i.e., only one source possesses the dominant energy at each time instant, does not hold for the two sets of source signals shown in Fig. 1. Therefore, it is necessary and interesting to develop new UBSS approaches which depend on milder constraints on the source signals and thus can be applied to a wider variety of applications.

III. THE PROPOSED UBSS ALGORITHM

In this section, we will develop a new group-based UBSS algorithm which exploits the statistical independence among different source signals to relax the sparsity constraint. Our developed UBSS algorithms are based on the following four conditions:

A2) Any M column vectors in the mixing matrix $\mathbf{A}$ are linearly independent.

A3) The source signals $s_1(k), s_2(k), \ldots, s_N(k)$ are statistically independent.

A4) At most $M - 1$ sources among N sources are active on any sample point.

A5) For any source signal $s_i(k)(1 \leq i \leq N)$, there exist some sample segments consisting of $M - 1$ adjacent sample points. On these sample segments, $s_i(k)$ and other $M - 2$ sources are active.

The assumption A2) is a mild restriction on the mixing matrix. It is easy to see that it can be satisfied for any randomly-generated mixing matrix $\mathbf{A}_{M \times N}$ with the probability one. The assumption A3) is also a widely-used assumption in the BSS field, which may be satisfied in many practical applications. Based on this assumption, a large number of ICA approaches have been developed in the past decades.

The assumption A4) is a mild sparsity constraint about source signals, which has been widely exploited in many existing works [19] solving the UBSS problem.

Next, let us examine whether the assumption A5) is easily satisfied. From the assumption A4), we know that at most $M - 1$ sources are active on any sample point, which means that active sources on any given sample point have C_N^{M-1} possible different combinations. Clearly, the probability that $s_i(k)$ and other $M - 2$ sources are active on a given sample point is $1/\mathrm{C}_N^{M-1}$. And then, the probability that active sources on $M - 1$ adjacent sample points consist of $s_i(k)$ and other $M - 2$ sources is equal to $\left(1/\mathrm{C}_N^{M-1}\right)^{M-1}$. In other words, the probability that the assumption A5) is met is equal to $\left(1/\mathrm{C}_N^{M-1}\right)^{M-1}$. Clearly, this assumption is not restrictive for practical application with thousands of sample points. In the application example shown in Fig. 1. with $M = 3$ mixture signals and $N = 4$ source signals, the probability that any sample segment with the length $M - 1$ satisfies the assumption A5) equals to 1/36, and 333 sample segments among total 12000 sample segments meet the assumption A5).

In summary, both the assumptions A4) and A5) are some mild sparsity constraints about source signals. They only require that source signals are partially sparse. In contrast, it is required by [16]–[18] that source signals are sparse on all sample points, and [15], [29]–[31] need some sample points or areas, which contain only one active source.

Since the sparsity and independence are two useful properties of source signals, clearly, it is interesting to combine these two properties to solve the UBSS problem, which is the motivation of this work. The combination of the sparsity and independence results in the following two advantages:

(1) Exploiting the sparsity assumption, we transform the original underdetermined BSS problem into a series of locally overdetermined problems by grouping the observed mixture signals, and make the classical ICA techniques, e.g., the JADE algorithm [35], can be extended to solve the underdetermined BSS problem.

(2) The introduction of independence can be used to relax the sparsity constraint to some extent.

At the same time, it is also notable that no constraint is imposed on the richness of the source signals, whilst [19] does. This makes the proposed algorithm be able to perform two UBSS problems shown in Fig. 1, which will be further illustrated in the simulation section, while [19] only can deal with blind separation of four source signals in the left column of Fig. 1.

The basic idea of the proposed approach is to transform the original UBSS problem into a series of non-underdetermined problems via grouping all mixture vectors $\mathbf{x}(k)$, $k = 1, 2, \ldots, K$ into multiple different sets, in each of which only $Q(Q < M)$ source signals are active. After this grouping step, what we need to deal with in each set is a non-underdertermined BSS problem with Q sources and $M(M > Q)$ mixtures. Thus, the conventional non-underdetermined blind identification algorithms, such as the JADE algorithm [35], can be applied to each set to estimate the steering vectors $\mathbf{a}_{\theta_1}, \mathbf{a}_{\theta_2}, \ldots, \mathbf{a}_{\theta_Q}$ associated with the Q active sources $s_{\theta_1}(k), s_{\theta_2}(k), \ldots, s_{\theta_Q}(k)$ in that set. Then, the estimate of the mixing matrix $\mathbf{A}$ can be computed by collecting and clustering the steering vectors obtained from all sets. Once the mixing matrix is estimated, the source signals can be recovered from their mixtures by exploiting the assumptions A2) and A4). The above idea is also used by the work of Karoui *et al.* in [34], which and the present paper share a common target, i.e., solving a globally underdetermined but locally (over)determined BSS problem. Their main differences lie in two aspects.

Firstly, in the stage of estimating the mixing matrix $\mathbf{A}$, [34] requires that there exist some single-source zones for each source, where only the considered source signal is active, and all other sources are inactive. On the contrary, our method only require that there are some sample segments for each source, where at most $M - 1$ sources including the considered source are active.

Secondly, during recovering sources after estimating the mixing matrix, [34] exploits the non-negativity property of remote sensing images to achieve the recovery of sources, while our proposed estimations for source signals depend on a sparsity constraint regarding source signals.

In the following, we shall present the new UBSS approach in detail, which consists of three main procedures.

A. GROUPING THE MIXTURE VECTORS

We first construct some $M \times (M - 1)$ mixture matrices $\mathbf{X}_p$ $(p = 1, 2, \cdots)$, each of which is composed of the samples of the mixture vector $\mathbf{x}(k)$ observed at $M - 1$ adjacent sample

points, i.e.,

$$\mathbf{X}_p = \left[\mathbf{x}(P_0) \ \mathbf{x}(P_0 + 1) \ \ldots \ \mathbf{x}(P_0 + M - 2) \right]_{M \times (M-1)} \quad (2)$$

where $P_0 = (p - 1) \times (M - 1) + 1$. Corresponding to the mixture matrix $\mathbf{X}_p$, a source matrix $\mathbf{S}_p$ is defined as follows:

$$
\mathbf{S}_p \\
= \begin{bmatrix}
s_{\theta_1}(P_0) & s_{\theta_1}(P_0+1) & \ldots & s_{\theta_1}(P_0+M-2) \\
s_{\theta_2}(P_0) & s_{\theta_2}(P_0+1) & \ldots & s_{\theta_2}(P_0+M-2) \\
\vdots & \vdots & \vdots & \vdots \\
s_{\theta_Q}(P_0) & s_{\theta_Q}(P_0+1) & \ldots & s_{\theta_Q}(P_0+M-2)
\end{bmatrix}_{Q \times (M-1)}
$$

$$(3)$$

where $s_{\theta_1}(k), s_{\theta_2}(k), \ldots, s_{\theta_Q}(k)$ are the active sources in the sample interval $[P_0, P_0 + M - 2]$, and Q stands for the number of these active sources.

Clearly, the assumption A5) results in $Q < M$ on some sample segments. In the following analysis, we will only consider those sample segments satisfying the assumption A5) for the time being. From (1), we have the following relation:

$$\mathbf{X}_p = \mathbf{A}_p \mathbf{S}_p \quad (4)$$

where

$$\mathbf{A}_p = \left[\mathbf{a}_{\theta_1}, \mathbf{a}_{\theta_2}, \ldots, \mathbf{a}_{\theta_Q} \right]_{M \times Q} \quad (5)$$

is composed of all the steering vectors associated with the active sources in the time interval $[P_0, P_0 + M - 2]$. In (4), additive noises are not considered for the convenience of analysis. Regarding the effects of noises, a detailed discussion will be provided in the next section.

Note that the source matrix $\mathbf{S}_p$ has Q rows and $M - 1$ columns with $Q \leq M - 1$, and its rows correspond to different source signals, which are mutually independent according to the assumption A3). Thus, we have the following lemma.

Lemma 1: The row vectors in the matrix $\mathbf{S}_p$ are linearly independent with the probability one.

On the other hand, due to the assumption A2) and $Q < M$, it is clear that the matrix $\mathbf{A}_p$ must be of full column rank, i.e.,

$$\text{rank} \left[\mathbf{A}_p \right] = Q. \quad (6)$$

According to Lemma 1, it holds from (4) and (6) that

$$\text{rank} \left[\mathbf{X}_p \right] = \text{rank} \left[\mathbf{A}_p \right] = Q \quad (7)$$

and

$$\text{span} \left\{ \mathbf{X}_p \right\} = \text{span} \left\{ \mathbf{A}_p \right\}. \quad (8)$$

From (7) and (8), we can see that the column vectors in any given mixture matrix $\mathbf{X}_p$ span a Q-dimensional subspace ($Q < M$). Clearly, the subspace clustering algorithms [27], [28] can be used to group these mixture matrices into different sets such that the mixture matrices in each set share the same Q-dimensional column space, which means that only Q source signals are active in these mixtures.

Next, we will provide a simpler scheme to group the mixture matrices into different sets. From (4) and (5), we can obtain the following theorem.

Theorem 1: Suppose that the sets $\widetilde{\mathcal{S}}_p$ and $\widetilde{\mathcal{S}}_q$ contain all the active sources in the mixture matrices $\mathbf{X}_p$ and $\mathbf{X}_q$, respectively. Then, it holds that $\widetilde{\mathcal{S}}_p$ is a subset of $\widetilde{\mathcal{S}}_q$ if and only if

$$\text{span} \left\{ \mathbf{X}_p \right\} \subseteq \text{span} \left\{ \mathbf{X}_q \right\}. \quad (9)$$

Proof: See Appendix A. ∎

Theorem 1 provides a theoretical foundation for grouping the mixture matrices into different sets such that each set has the same active sources. The process of obtaining one such set is as follows. First, for a given mixture matrix $\mathbf{X}_q$, find all the mixture matrices $\mathbf{X}_{p_i}(i = 1, 2, \ldots,)$ satisfying (9). Then, construct the set $\mathcal{G}_q$ by $\mathcal{G}_q = \left\{ \mathbf{X}_q, \mathbf{X}_{p_1}, \mathbf{X}_{p_2}, \ldots \right\}$. Based on Theorem 1 and the assumption A4), it is clear that the number of active sources in the mixture set $\mathcal{G}_q$ must be less than the number of mixtures, i.e., $Q < M$. This implies that in the mixture set $\mathcal{G}_q$, we only need to solve a non-underdermined blind problem.

It is notable that the above theoretical analysis does not consider the existence of noises. Next, let us consider how to derive an effective approach to group the mixture matrices in noisy environments. From (1), (4) has the following form in noisy environments:

$$\mathbf{X}_p = \mathbf{A}_p \mathbf{S}_p + \mathbf{N}_p \quad (10)$$

where

$$\mathbf{N}_p = \left[\mathbf{n}(P_0) \ \mathbf{n}(P_0 + 1) \ \ldots \ \mathbf{n}(P_0 + M - 2) \right]. \quad (11)$$

Clearly, due to the existence of noises $\mathbf{n}(k)$, all mixture matrices $\mathbf{X}_p$ defined in (10) must be of full column rank. Considering that the dimension of the matrix $\mathbf{X}_p$ is $M \times (M - 1)$, we can easily see that the orthogonal complement $\mathbf{v}_p$ of column spaces of $\mathbf{X}_p$ is a one-dimensional subspace. From the above observations, it is easy to see that if $\mathbf{X}_p$ and $\mathbf{X}_q$ in (10) have the same set of active sources, that is, $\mathbf{A}_p = \mathbf{A}_q$, then their orthogonal complements $\mathbf{v}_p$ and $\mathbf{v}_q$ will be close to each other in noisy environment with high or moderate SNRs. The above conclusion can be summarized as the following Proposition 1, which provides a theoretical foundation for developing an effective approach to group the mixture matrices in noise environment.

Proposition 1: Suppose that the vector $\mathbf{v}_p$ is the orthogonal complement of column space of the mixture matrix $\mathbf{X}_p$. It holds that $\mathbf{X}_p$ and $\mathbf{X}_q$ have the same set of active sources, if

$$\frac{\| \mathbf{v}_p^T \cdot \mathbf{X}_q \|}{\| \mathbf{v}_p \| \cdot \| \mathbf{X}_q \|_F} < \varepsilon \quad (12)$$

where ε is a positive and small threshold.

On the other hand, let us consider those sample segments violating the assumption A5). Clearly, for the mixture matrix $\mathbf{X}_p$ associated with these sample segments, few other matrices $\mathbf{X}_q$ can satisfy the condition in (12). Based on the above conclusions, we propose a detailed procedure to group the mixture matrices, which is shown in Table I. Through this grouping procedure, we can obtain the sets $\mathcal{G}_1, \mathcal{G}_2, \ldots, \mathcal{G}_J$, each of which consists of a number of mixture matrices containing $Q(Q < M)$ identical active sources.

TABLE 1. Procedure of grouping mixture matrices.

---

(1) Initialize the set $\mathcal{U}$ as $\mathcal{U} = \{\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_p, \ldots\}$ which contains all mixture matrices. And set $J = 1$.

(2) Construct an empty set $\mathcal{G}_J$.

(3) Select randomly a matrix $\mathbf{X}_j$ from the set $\mathcal{U}$.

(4) Execute the singular value decomposition of $\mathbf{X}_j$, and let the unit vector $\mathbf{v}_j$ be the orthogonal complement of column space of $\mathbf{X}_j$.

(5) Move $\mathbf{X}_j$ and all the matrices $\mathbf{X}_i$ satisfying $\dfrac{\|\mathbf{v}_i^T \cdot \mathbf{X}_j\|}{\|\mathbf{v}_i\| \cdot \|\mathbf{X}_j\|_F} < \varepsilon$ from the set $\mathcal{U}$ into the set $\mathcal{G}_J$. If the number of elements in $\mathcal{G}_J$ is smaller than the threshold D, then delete all the elements from $\mathcal{G}_J$, and return to the step (3).

(6) If $\mathcal{U}$ becomes an empty set, the algorithm stops and we obtain some sets $\mathcal{G}_1, \mathcal{G}_2, \ldots, \mathcal{G}_J$. If not, set $J = J+1$ and go to the step (2).

B. IDENTIFYING THE MIXING MATRIX

Since in each set $\mathcal{G}_j$, the number of active sources is less than that of mixtures (namely, $Q < M$), we only need to solve a non-underdetermined blind problem. Thus, the existing non-underdetermined blind identification algorithms, e.g., the JADE algorithm [35], can be employed to estimate the steering vectors associated with all active sources in each set $\mathcal{G}_j(j = 1, \ldots, J)$, respectively. For example, suppose that $s_{j_1}(k), s_{j_2}(k), \ldots, s_{j_Q}(k)$ are the active sources in the set $\mathcal{G}_j$, and their corresponding steering vectors are $\mathbf{a}_{j_1}, \mathbf{a}_{j_2}, \ldots, \mathbf{a}_{j_Q}$, respectively. Clearly, we can obtain the estimates of these steering vectors from the mixture matrices contained in the set $\mathcal{G}_j$ by using an existing blind identification algorithm. In order to avoid the gain ambiguity in estimating the steering vectors, we normalize all the estimated steering vectors to a unit sphere, and make their first elements be non-negative.

After estimating the steering vectors associated with all active sources in each set $\mathcal{G}_1, \mathcal{G}_2, \ldots, \mathcal{G}_J$, we can cluster all these estimated steering vectors into N groups. Then, we take the centers of these groups as the column vectors of the estimation $\widetilde{\mathbf{A}}$ for the mixing matrix $\mathbf{A}$. To this point, the estimation of the mixing matrix $\mathbf{A}$ is completed.

C. RECOVERING THE SOURCE SIGNALS

Although the mixing matrix $\mathbf{A}$ has been estimated, recovering the source signals is not a trivial task in the underdetermined case as the mixing matrix is not invertible. In order to achieve source recovery, we need the assumption A4), and proceed with the definition of a set $\mathcal{T}$ as follows.

Definition 1: $\mathcal{T}$ is a set composed of all $M \times (M - 1)$ submatrices of the matrix $\mathbf{A}$, i.e.,

$$\mathcal{T} = \left\{ \mathcal{T}^{(i)} \middle| \mathcal{T}^{(i)} = \left[\mathbf{a}_{\theta_1}, \mathbf{a}_{\theta_2}, \ldots, \mathbf{a}_{\theta_{(M-1)}} \right] \right\}. \quad (13)$$

Clearly, $\mathcal{T}$ contains C_N^{M-1} matrix elements. With respect to the set $\mathcal{T}$, we have the following conclusion.

Theorem 2: Under the assumption A4), for any given mixture vector $\mathbf{x}(k)$, there must exist a matrix element $\mathcal{T}^{(*)} = \left[\mathbf{a}_{\alpha_1}, \mathbf{a}_{\alpha_2}, \ldots, \mathbf{a}_{\alpha_{M-1}} \right]$ in the set $\mathcal{T}$, such that

$$\left[\mathbf{I}_{M \times M} - \mathcal{T}^{(*)} \mathrm{pinv}\left(\mathcal{T}^{(*)} \right) \right] \cdot \mathbf{x}(k) = \mathbf{0}_{M \times 1}. \quad (14)$$

Proof: See Appendix B. ∎

After finding the matrix element $\mathcal{T}^{(*)}$ satisfying (14) from the set $\mathcal{T}$, let us construct an N-dimensional vector $\hat{\mathbf{s}}(k)$, which is built by making its α_ith element equal the ith element of $\mathrm{pinv}\left(\mathcal{T}^{(*)} \right) \mathbf{x}(k)$ for $i = 1, 2, \ldots, M - 1$, and its all other components be zero. Then, based on Theorem 2, we propose another theorem as follows.

Theorem 3: Under the assumption A4), for any time instant k, if in the set $\mathcal{T}$, the matrix element $\mathcal{T}^{(*)}$ satisfies (14), then the N-dimensional vector $\hat{\mathbf{s}}(k)$ is equal to the source signal vector $\mathbf{s}(k)$ with probability one. ∎

Proof: See Appendix C. ∎

Theorem 3 clearly shows that the matrix element $\mathcal{T}^{(*)}$ satisfying (14) can be used to find the estimate of the original source signal vector $\mathbf{s}(k)$. This paves the way for recovering $\mathbf{s}(k)$ from the observed mixture vector $\mathbf{x}(k)$ by exploiting the estimated mixing matrix $\mathbf{A}$. It should be noted that in a noisy environment, the element $\mathcal{T}^{(*)}$ satisfying (14) can be found by the following criterion:

$$\mathcal{T}^{(*)} = \arg \min_{\mathcal{T}^{(l)} \in \mathcal{T}} \left\| \left[\mathbf{I}_{M \times M} - \mathcal{T}^{(l)} \mathrm{pinv}\left(\mathcal{T}^{(l)} \right) \right] \mathbf{x}(k) \right\|^2. \quad (15)$$

In summary, our proposed UBSS algorithm is formulated as follows:

- Step 1. Construct a series of mixture matrices $\mathbf{X}_p$ ($p = 1, 2, \cdots$) by

$$\mathbf{X}_p = [\mathbf{x}(P_0), \mathbf{x}(P_0 + 1), \ldots, \mathbf{x}(P_0 + M - 2)]_{M \times (M-1)}$$

  where $P_0 = (p - 1) \times (M - 1) + 1$.
- Step 2. Use the procedure provided in Table I to group all mixture matrices $\mathbf{X}_p$ ($p = 1, 2, \cdots$) into different sets $\mathcal{G}_1, \mathcal{G}_2, \ldots, \mathcal{G}_J$.
- Step 3. In each set $\mathcal{G}_j(j = 1, \ldots, J)$, use a blind identification approach, e.g., the JADE algorithm [35], to estimate the steering vectors associated with the active sources in $\mathcal{G}_j$.
- Step 4. Normalize all of the estimated steering vectors and ensure that the first element of each normalized steering vector is nonnegative.
- Step 5. Cluster the normalized steering vectors obtained in all sets $\mathcal{G}_j(j = 1, \ldots, J)$ into N groups. Then, take the centers of these groups as the column vectors of the estimation $\widetilde{\mathbf{A}}$ for the mixing matrix $\mathbf{A}$.
- Step 6. Utilize the estimate of the mixing matrix $\mathbf{A}$ to construct the set $\mathcal{T}$ according to Definition 1.
- Step 7. For each mixture vector $\mathbf{x}(k)$ ($k = 1, 2, \ldots, K$), find the matrix element $\mathcal{T}^{(*)}$ satisfying (15) from the set $\mathcal{T}$.

- Step 8. Suppose that $\mathcal{T}^{(*)}$ contains the α_ith column of the matrix $\mathbf{A}$, $i = 1, 2, \ldots, M - 1$, and the $(M - 1)$-dimensional vector $\mathbf{r}(k) = [r_1, r_2, \ldots, r_{M-1}]^T = \text{pinv}\left(\mathcal{T}^{(*)}\right) \mathbf{x}(k)$. Then the source signal vector can be estimated as $\hat{\mathbf{s}}(k) = \left[\hat{s}_1, \hat{s}_2, \ldots, \hat{s}_N\right]^T$, in which $\hat{s}_{\alpha_i} = r_i$(for any $i = 1, 2, \ldots, M - 1$) and $\hat{s}_i = 0$(for any $i \notin \{\alpha_1, \alpha_2, \ldots, \alpha_{M-1}\}$).

Discussions about the problem of permutation and gain ambiguities: It is well-known that the solution for the BSS problem suffers from permutation and gain ambiguities. Next, we will explain in detail that the effectiveness of our proposed algorithm is not affected by the above problems.

As shown in Step 2, our proposed algorithm group firstly sample points. After then, Step 3 uses the JADE algorithm to estimate the mixture matrix in each set. Clearly, each column vector in these estimated matrices is identical to some column vector in the true mixture matrix $\mathbf{A}$ with different gains. In order to remove the ambiguity with these different scale gains and negative gains, we normalize all of these estimated column vectors from all sets to the unit sphere, and make the first element of these vectors be nonnegative, as shown in Step 4. Next, Step 5 clusters these column vectors to obtain an estimation $\widetilde{\mathbf{A}}$ for the true mixture matrix $\mathbf{A}$. Finally, exploiting this estimated mixture matrix $\widetilde{\mathbf{A}}$ and the assumption A4), we can recover source signals in each sample point, as shown in Step A6, Step A7 and Step A8.

In summary, in each set, we only estimate some column vectors of the mixture matrix, and source recovery is not considered for the time being. After normalizing all these estimated column vectors, we cluster these vectors to obtain the estimation $\widetilde{\mathbf{A}}$ for the true mixture matrix $\mathbf{A}$. Finally, source signals are recovered by using the estimated mixture matrix $\widetilde{\mathbf{A}}$. From the above analysis, it is easy to find that the permutation and gain ambiguities cannot affect the effectiveness of the proposed algorithm.

IV. SIMULATIONS

In this section, we will use some simulation examples to illustrate the effectiveness of the proposed algorithm. In all examples, the JADE algorithm [35] will be used to solve locally non-underdetermined problem in each set.

Example 1: In the first simulation, we will consider to use our proposed algorithm to achieve blind identification of a underdetermined mixture system with $M = 3$ outputs and $N = 4$ inputs. In this system, the 3×4 mixing matrix is given by:

$$\mathbf{A} = \begin{bmatrix} 6.1560 & -5.3550 & 6.2050 & 4.9960 \\ 5.7850 & 7.0110 & 5.5270 & -6.1030 \\ 6.0020 & 6.2240 & -6.1770 & 5.6450 \end{bmatrix}, \quad (16)$$

and 4 source signals are taken from four speech segments, whose waveforms are shown in Fig. 2. The parameter ε in Table I is taken as $\varepsilon = 0.12$. By using from Step 1 to Step 5 in our proposed algorithm, we can obtain the estimation $\widetilde{\mathbf{A}}$ for
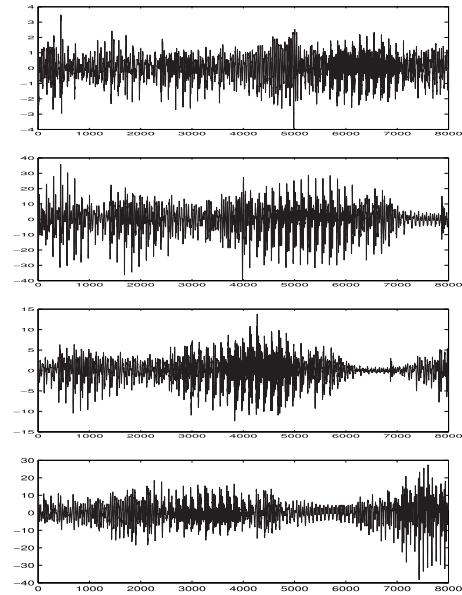

FIGURE 2. Speech source signals.

the mixing matrix $\mathbf{A}$. After removing the order indeterminacy and normalizing each column in $\widetilde{\mathbf{A}}$ and $\mathbf{A}$, we compute $\widetilde{\mathbf{A}}^T \mathbf{A}$ to obtain the following results:

$$\widetilde{\mathbf{A}}^T \mathbf{A} = \begin{bmatrix} 0.9913 & 0.4608 & 0.1809 & 0.3485 \\ 0.3638 & 0.9692 & -0.0945 & -0.5327 \\ 0.1383 & -0.4207 & 0.9835 & -0.4004 \\ 0.2734 & -0.1832 & -0.5262 & 0.9824 \end{bmatrix}. \quad (17)$$

It is easy to see that all diagonal elements in the matrix $\widetilde{\mathbf{A}}^T \mathbf{A}$ is very closed to 1, which means that the i-th$(i = 1, 2, 3, 4)$ column in $\widetilde{\mathbf{A}}$ is a good estimation for the corresponding column in $\mathbf{A}$.

Next, let us consider another set of four source signals, each of which consist of 1200 samples, The waveforms of these four random-generated source signals are illustrated in Fig. 3. We can achieve the task of blind channel identification by using from Step 1 to Step 5 in our proposed algorithm, and obtain the following estimation for the mixture matrix after removing the indeterminacy of order and scale:

$$\widetilde{\mathbf{A}} = \begin{bmatrix} 5.1149 & -6.0321 & 6.1838 & 5.3403 \\ 6.5224 & 6.2444 & 5.9155 & -5.7987 \\ 6.0867 & 6.3349 & -5.8037 & 5.6318 \end{bmatrix}. \quad (18)$$

In this simulation test, the threshold ε is taken as 0.01. We use the index $\theta(\mathbf{a}, \mathbf{b})$ to evaluate the distance between two vectors $\mathbf{a}$ and $\mathbf{b}$, which is given by

$$\theta(\mathbf{a}, \mathbf{b}) = 1 - \frac{\left|\mathbf{a}^T \mathbf{b}\right|}{\|\mathbf{a}\| \cdot \|\mathbf{b}\|}. \quad (19)$$

Clearly, the smaller the index $\theta(\mathbf{a}, \mathbf{b})$ is, the closer the vector $\mathbf{a}$ lies to $\mathbf{b}$. With regard to column vectors in the mixing matrix $\mathbf{A}$ in (16) and the estimated matrix $\widetilde{\mathbf{A}}$ in (18), we have
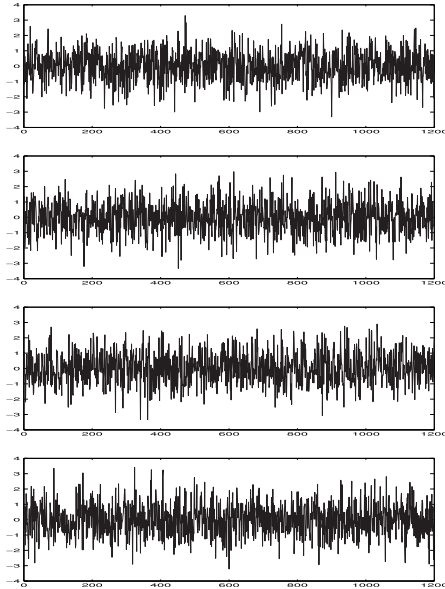
FIGURE 3. Randomly generated source signals.

FIGURE 4. Performance comparison among the proposed algorithm, FOBIUM Algorithm and TIFROM Algorithm.

FIGURE 5. Simulation result in separating the four source signals shown in the right column of Fig. 1 from their three mixtures. First row: source signals. Second row: mixture signals. Third row: recovered source signals.

the following results for the performance index $\theta(\mathbf{a}, \mathbf{b})$:

$$\theta(\mathbf{A}(:, 1), \tilde{\mathbf{A}}(:, 1)) = 0.0076, \tag{20}$$

$$\theta(\mathbf{A}(:, 2), \tilde{\mathbf{A}}(:, 2)) = 0.0046, \tag{21}$$

$$\theta(\mathbf{A}(:, 3), \tilde{\mathbf{A}}(:, 3)) = 0.0014, \tag{22}$$

$$\theta(\mathbf{A}(:, 4), \tilde{\mathbf{A}}(:, 4)) = 0.0011. \tag{23}$$

The above simulation results show that our proposed algorithm can obtain satisfactory channel identification performance for randomly-generated source signals.

In order to evaluate further the performance of channel identification of the proposed algorithm, we compare this algorithm with FOBIUM algorithm and TIFROM algorithm. FOBIUM algorithm is a non-sparsity UBSS algorithm based on fourth-order cumulant [5], [36], and TIFROM algorithm is a classical time-frequency representations-based UBSS algorithm [31]. In this example, we use the following performance index

$$\text{error} = \frac{1}{N} \cdot \sum_{i=1}^{N} \theta(\mathbf{A}(:, i), \tilde{\mathbf{A}}(:, i)) \tag{24}$$

where N denotes the number of columns in the mixing matrix. Fig. 4. shows the above performance indexes of these three algorithms over different noise levels, which are evaluated by 100 Monte Carlo runs. From the simulation results shown in Fig. 4., it is seen that our proposed algorithm can achieve satisfactory performance for channel identification.

Example 2: Let us consider a UBSS problem with $M = 3$ mixtures, and $N = 4$ source signals $t_1(k), t_2(k), t_3(k), t_4(k)$ whose waveforms are shown in the right column in Fig. 1. As discussed in Section II, since these four source signals do not satisfy the assumption A1) and the assumption presented in [16]–[18] that only one source is active at each time instant, UBSS cannot be achieved by the methods
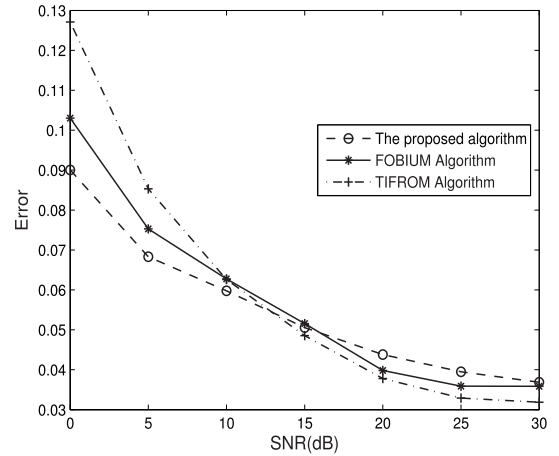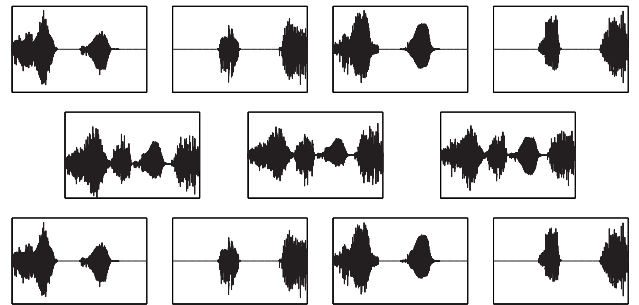
in [16]–[18] and [19]. In contrast, the proposed UBSS algorithm can separate the source signals $t_1(k), t_2(k), t_3(k), t_4(k)$ from their underdetermined mixtures as it does not rely on the assumption A1) and the restrictive sparsity constraint required by [16]–[18]. Fig. 5 illustrates the separation result by our algorithm, where the source signals, the observed mixture signals and the recovered source signals are shown in the first, second and third rows, respectively.

Next, in order to further illustrate the effectiveness of the proposed UBSS algorithm, we consider another UBSS task with four source signals shown in the first row of Fig. 6. The randomly-generated mixing matrix is of dimension 3×4 and three observed mixing signal are shown in the second row of Fig. 6. The proposed UBSS algorithm is used to separate these four source signals from their three mixtures. The recovered source signals are displayed in the third row of Fig. 6. From the above two simulation results, we can see that the proposed algorithm has achieved satisfactory UBSS performance.

Example 3: In this example, Monte Carlo runs are used to evaluate the performance of the proposed algorithm versus signal to noise ratio (SNR). The mean squared error (MSE) is utilized as the performance index, which is defined as
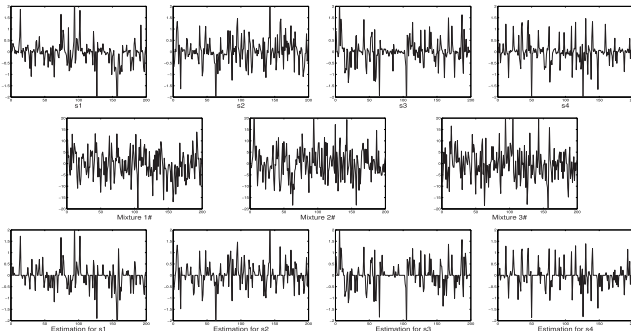
FIGURE 6. Simulation result in separating randomly-generated source signals. First row: source signals. Second row: mixture signals. Third row: recovered source signals.
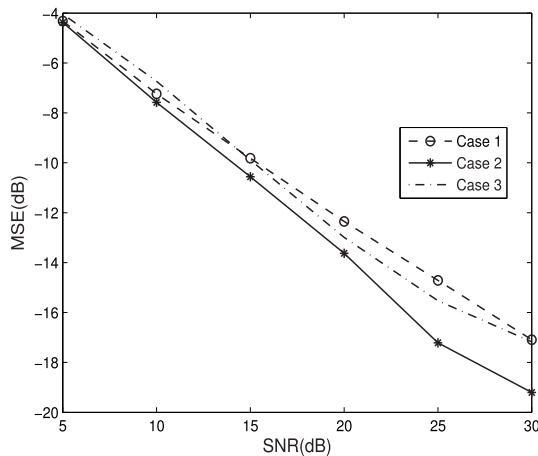

FIGURE 7. Performance comparison under different source signals.

FIGURE 8. Performance comparison between the proposed algorithm and the algorithm in [19].

follows [10]:

$$\text{MSE(dB)} = 10\log_{10}\left(\frac{\min_\delta \sum_{k=1}^{K} |\delta\hat{s}(k) - s(k)|^2}{\sum_{k=1}^{K} |s(k)|^2}\right) \quad (25)$$

where $\hat{s}(k)$ is the estimate of the source signal $s(k)$, δ is a scalar reflecting the scalar ambiguity, and K denotes the number of samples. In this example, additive noise signals with the normal distribution are considered. The parameter ε in Table I is taken as $\varepsilon = 0.2$. Fig. 7. shows the performance of the proposed algorithm evaluated over 100 Monte Carlo runs in the following three scenarios:

Case 1: The source signals are $s_1(k), s_2(k), s_3(k), s_4(k)$ shown in the left column of Fig. 1.

Case 2: The source signals are $t_1(k), t_2(k), t_3(k), t_4(k)$ shown in the right column of Fig. 1.

Case 3: The source signals are the four randomly generated signals shown in the first row of Fig 6.

It can be seen from Fig. 7 that as expected, MSE decreases with the increase of SNR in all three cases. The proposed algorithm achieves satisfactory separation performance.

Next, we compare the performance of the proposed algorithm with that of the UBSS algorithm in [19]. For this purpose, we use $s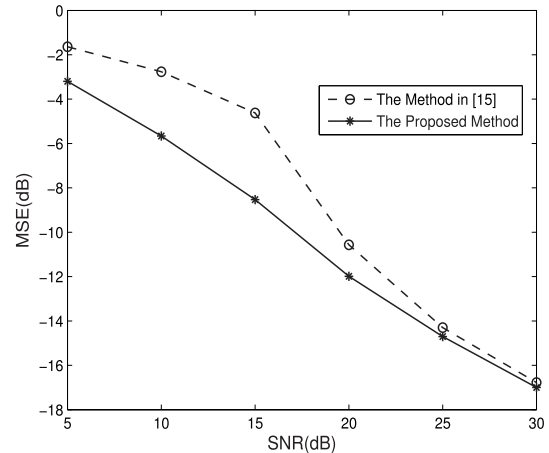_1(k), s_2(k), s_3(k), s_4(k)$ shown in the left column of Fig. 1 as source signals since they satisfy the assumptions about source signals required by the two algorithms. Fig. 8 shows the performances of the two algorithms under different SNRs. We can see that under lower SNRs, our algorithm outperforms the one in [19] with large margins. This is because the proposed algorithm exploits both the statistical property and the sparsity of the source signals, whilst the algorithm in [19] only uses the sparsity property to estimate the mixing matrix and the source signals. With the rise of SNR, the performance gap between the two algorithms reduces gradually.

Example 4: In this example, we will consider a more real application environment, where some microphones are placed on different positions to receive the mixed audio signals. In the simulation, a number of computers are used to play different sound files at the same time for 90 seconds. Our proposed algorithm is used to separate these source audio signals from the mixture signals received by microphones, and the performance of source recovery is evaluated by using MSE(dB) defined in (25). The above simulation test is executed four times in different scenarios:

Scenario 1: 3 microphones and 5 source signals;
Scenario 2: 3 microphones and 6 source signals;
Scenario 3: 2 microphones and 3 source signals;
Scenario 4: 2 microphones and 4 source signals.

TABLE 2. Performance comparisons of the proposed UBSS algorithm under different scenarios.

Scenarios	MSE(dB)
3 **microphones and** 5 **source signals**	−15.85
3 **microphones and** 6 **source signals**	−14.75
2 **microphones and** 3 **source signals**	−18.43
2 **microphones and** 4 **source signals**	−17.11

Performance of the proposed algorithm with different scenarios is shown in Table II. It is easy to see that the

proposed algorithm can achieve satisfactory performance under different scenarios, which has a close relationship with the difference between the number of source signals and that of mixture signals.

V. CONCLUSION

This paper addresses the problem of underdetermined blind separation of N sources from their M instantaneous mixtures ($N > M$). By combining both sparsity and independence of source signals, the proposed UBSS approach transforms the original underdetermined problem into a series of overdetermined BSS problems. This object is achieved by grouping all the observed mixture vectors into a number of different sets such that each set only contains $Q(Q < M)$ active sources. In each set, we only need to solve an overdetermined blind source separation problem with M outputs and $Q(Q < M)$ mutually independent inputs, which can be easily tackled by some ICA techniques.

By exploiting the statistical independence of sources, this work relaxes the sparsity restriction about sources to some extent. At the same time, our UBSS approach does not impose any richness constraint on sources, which is required by [19]. The effectiveness of the new approach is illustrated by theoretical analysis and simulation results.

APPENDIX A
PROOF OF THEOREM 1

Proof of Necessity: From (4) and (5), it is easy to find that the necessity must hold.

Proof of Sufficiency: Assume that the sufficiency does not hold, i.e., when span $\{\mathbf{X}_p\} \subseteq$ span $\{\mathbf{X}_q\}$, $\widetilde{\mathcal{S}}_p$ is not a subset of $\widetilde{\mathcal{S}}_q$. This means that there exists at least one source signal $s^*(k)$ satisfying $s^*(k) \in \widetilde{\mathcal{S}}_p$ and $s^*(k) \notin \widetilde{\mathcal{S}}_q$. Thus, the steering vector $\mathbf{a}^*$ associated with $s^*(k)$ is not a column vector in the matrix $\mathbf{A}_q$ but a column vector in the matrix $\mathbf{A}_p$, where $\mathbf{A}_p$ and $\mathbf{A}_q$ consist of all the steering vectors associated with the active sources in the sets $\widetilde{\mathcal{S}}_p$ and $\widetilde{\mathcal{S}}_q$, respectively. From (8), we obtain

$$\text{span}\{\mathbf{X}_p\} = \text{span}\{\mathbf{A}_p\} \quad \text{and} \quad \text{span}\{\mathbf{X}_q\} = \text{span}\{\mathbf{A}_q\}. \quad (26)$$

Considering that $\mathbf{a}^*$ is a column vector in the matrix $\mathbf{A}_p$, it follows from (26) that

$$\mathbf{a}^* \in \text{span}\{\mathbf{A}_p\} = \text{span}\{\mathbf{X}_p\} \subseteq \text{span}\{\mathbf{X}_q\} = \text{span}\{\mathbf{A}_q\}. \quad (27)$$

Since $\mathbf{a}^*$ is not a column vector in the matrix $\mathbf{A}_q$, $\mathbf{a}^* \in$ span $\{\mathbf{A}_q\}$ in (27) means that $\mathbf{a}^*$ and all the Q column vectors in $\mathbf{A}_q$ are linearly dependent. Due $Q \leq M - 1$, clearly, the above conclusion contradicts with the assumption A2). This completes the proof.

APPENDIX B
PROOF OF THEOREM 2

According to the assumption A4), at any given sample point k, the number of active sources must be less than the number of mixtures, i.e., $Q < M$. This means that the source signal vector $\mathbf{s}(k)$ contains at most $M - 1$ nonzero elements.

From the definition of the set $\mathcal{T}$, it is easy to see that there exists a matrix element $\mathcal{T}^{(*)}$ in $\mathcal{T}$ satisfying

$$\mathbf{As}(k) \in \text{span}\left\{\mathcal{T}^{(*)}\right\} \quad (28)$$

or equivalently,

$$\mathbf{As}(k) = \mathcal{T}^{(*)}\mathbf{s}' \quad (29)$$

where $\mathbf{s}'$ is an $(M - 1)$-dimensional column vector. From (1) and (29), it follows

$$\begin{aligned}
\left[\mathbf{I}_{M \times M} - \mathcal{T}^{(*)}\text{pinv}\left(\mathcal{T}^{(*)}\right)\right] &\mathbf{x}(k) \\
&= \mathbf{x}(k) - \mathcal{T}^{(*)}\text{pinv}\left(\mathcal{T}^{(*)}\right)\mathbf{As}(k) \\
&= \mathbf{x}(k) - \mathcal{T}^{(*)}\text{pinv}\left(\mathcal{T}^{(*)}\right)\mathcal{T}^{(*)}\mathbf{s}' \\
&= \mathbf{As}(k) - \mathcal{T}^{(*)}\mathbf{s}' \\
&= \mathbf{0}_{M \times 1}.
\end{aligned} \quad (30)$$

This completes the proof.

APPENDIX C
PROOF OF THEOREM 3

Given any sample point k, suppose that $\mathbf{r}(k)$ is an $(M - 1)$-dimensional column vector given by

$$\mathbf{r}(k) = \text{pinv}\left(\mathcal{T}^{(*)}\right)\mathbf{x}(k) = [r_1, r_2, \ldots, r_{M-1}]^T. \quad (31)$$

Then, the vector $\hat{\mathbf{s}}(k)$ can be expressed as

$$\begin{aligned}
\hat{\mathbf{s}}(k) &= \left[\hat{s}_1, \ldots, \hat{s}_{\alpha_1}, \ldots, \hat{s}_{\alpha_2}, \ldots, \hat{s}_{\alpha_{M-1}}, \ldots, \hat{s}_N\right]^T \\
&= [0, \ldots, r_1, \ldots, r_2, \ldots, r_{M-1}, \ldots, 0]. \quad (32)
\end{aligned}$$

From the assumption A4), at any given sample point k, the number of active sources must be less than the number of mixtures, i.e., $Q < M$. This means that the number of nonzero elements in the vector $\mathbf{s}(k)$ is less than M. Thus, we have from (1) and Definition 1 that there must exist a matrix element $\mathcal{T}^{(l)}$ in the set $\mathcal{T}$ such that

$$\mathbf{x}(k) = \mathbf{As}(k) \in \text{span}\left\{\mathcal{T}^{(l)}\right\}, \quad 1 \leq l \leq \mathrm{C}_N^{M-1}. \quad (33)$$

Clearly, all the mixture vectors $\mathbf{x}(k) = \mathbf{As}(k)$, $k = 1, 2, \ldots, K$ must be contained in the set $\mathcal{X}$ defined below:

$$\mathcal{X} = \bigcup_{l=1}^{\mathrm{C}_N^{M-1}} \text{span}\left\{\mathcal{T}^{(l)}\right\}.$$

Let $\mathcal{X}_0$ be the union of all the intersections of any two different subspaces span $\{\mathcal{T}^{(i)}\}$ and span $\{\mathcal{T}^{(j)}\}$, i.e.,

$$\mathcal{X}_0 = \bigcup_{i,j=1(i \neq j)}^{\mathrm{C}_N^{M-1}} \left[\text{span}\left\{\mathcal{T}^{(i)}\right\} \bigcap \text{span}\left\{\mathcal{T}^{(j)}\right\}\right]. \quad (34)$$

It is clear that $\mathcal{X}_0$ has a measure zero in $\mathcal{X}$, which means that any mixture vector $\mathbf{x}(k)$ is in the set $\mathcal{X}_0$ with probability zero.

Moreover, according to Theorem 2, one can find a matrix element $\mathcal{T}^{(*)}$ satisfying (14) in the set $\mathcal{T}$. Then, from (31), we have

$$
\begin{aligned}
\mathbf{x}(k) &= \mathcal{T}^{(*)}\text{pinv}\left(\mathcal{T}^{(*)}\right)\mathbf{x}(k) \\
&= \mathcal{T}^{(*)}\mathbf{r}(k).
\end{aligned} \tag{35}
$$

It holds from (32) and (35) that

$$
\begin{aligned}
\mathbf{A}\hat{\mathbf{s}}(k) &= \left[\mathbf{a}_{\alpha_1}, \mathbf{a}_{\alpha_2}, \ldots, \mathbf{a}_{\alpha_{M-1}}\right]\cdot\mathbf{r}(k) \\
&= \mathcal{T}^{(*)}\mathbf{r}(k) \\
&= \mathbf{x}(k).
\end{aligned} \tag{36}
$$

Furthermore, from (33), there must exist an $(M-1)$-dimensional vector $\bar{\mathbf{s}}(k)$ such that

$$
\mathbf{x}(k) = \mathbf{A}\mathbf{s}(k) = \mathcal{T}^{(l)}\bar{\mathbf{s}}(k). \tag{37}
$$

Based on (36) and (37), we obtain

$$
\mathbf{A}\left[\mathbf{s}(k) - \hat{\mathbf{s}}(k)\right] = \mathbf{0}. \tag{38}
$$

According to the number of nonzero elements in the vector $\mathbf{s}(k) - \hat{\mathbf{s}}(k)$, denoted by L, the following three cases are considered.

Case 1: $M < L \leq N$

By combining (36) and (37), we have

$$
\begin{cases}
\mathbf{x}(k) = \mathbf{A}\hat{\mathbf{s}}(k) = \mathcal{T}^{(*)}\mathbf{r}(k) \\
\mathbf{x}(k) = \mathbf{A}\mathbf{s}(k) = \mathcal{T}^{(l)}\bar{\mathbf{s}}(k).
\end{cases} \tag{39}
$$

If $\mathcal{T}^{(*)} = \mathcal{T}^{(l)}$ in (39), then the indices of all nonzero elements in the vectors $\mathbf{s}(k)$ and $\hat{\mathbf{s}}(k)$ are included in the set $\{\alpha_1, \alpha_2, \ldots, \alpha_{M-1}\}$ due to $\mathcal{T}^{(*)} = \mathcal{T}^{(l)} = \left[\mathbf{a}_{\alpha_1}, \mathbf{a}_{\alpha_2}, \ldots, \mathbf{a}_{\alpha_{M-1}}\right]$. Thus, the number of nonzero elements in the vector $\mathbf{s}(k) - \hat{\mathbf{s}}(k)$ cannot exceed $M-1$, i.e., $L \leq M-1$, which contradicts with $M < L \leq N$. On the other hand, if $\mathcal{T}^{(*)} \neq \mathcal{T}^{(l)}$ in (39), then the vector $\mathbf{x}(k)$ is included in the set $\mathcal{X}_0$ given in (34). Since $\mathcal{X}_0$ has a measure zero in $\mathcal{X}$, this event happens with probability zero.

Case 2: $1 \leq L \leq M$

In this case, the vector $\mathbf{s}(k) - \hat{\mathbf{s}}(k)$ contains L nonzero elements. From (38), it follows that there exist L linearly dependent column vectors in the matrix $\mathbf{A}$. Since $1 \leq L \leq M$, this conclusion obviously contradicts with the assumption A2).

Case 3: $L = 0$

If $L = 0$, it means that $\mathbf{s}(k) - \hat{\mathbf{s}}(k)$ is a zero vector, i.e., $\mathbf{s}(k) = \hat{\mathbf{s}}(k)$.

Based on the analysis outcomes from the above three cases, we can conclude that $\hat{\mathbf{s}}(k)$ must be equal to the source signal vector $\mathbf{s}(k)$ with probability one. This completes the proof.

REFERENCES

[1] C. Wei, W. L. Woo, and S. S. Dlay, "Nonlinear underdetermined blind signal separation using Bayesian neural network approach," *Digit. Signal Process.*, vol. 17, no. 1, pp. 50–68, Jan. 2007.

[2] Y. Xiang, W. Yu, J. Zhang, and S. An, "Blind source separation based on phase and frequency redundancy of cyclostationary signals," *IEICE Trans. Fundam.*, vol. E87-A, no. 12, pp. 3343–3349, Dec. 2004.

[3] V. Zarzoso, "Second-order criterion for blind source extraction," *Electron. Lett.*, vol. 44, no. 22, pp. 1327–1328, Oct. 2008.

[4] A. Ferreol, L. Albera, and P. Chevalier, "Fourth-order blind identification of underdetermined mixtures of sources (FOBIUM)," *IEEE Trans. Signal Process.*, vol. 53, no. 5, pp. 1640–1653, May 2005.

[5] L. D. Lathauwer, J. Castaing, and J. F. Cardoso, "Fourth-order cumulant-based blind identification of underdetermined mixtures," *IEEE Trans. Signal Process.*, vol. 55, no. 6, pp. 2965–2973, Jun. 2007.

[6] A. T. Cemgil, C. Févotte, and S. J. Godsill, "Variational and stochastic inference for Bayesian source separation," *Digit. Signal Process.*, vol. 17, no. 5, pp. 891–913, Sep. 2007.

[7] Y. Xiang, V. K. Nguyen, and N. Gu, "Blind equalization of nonirreducible systems using the CM criterion," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 53, no. 8, pp. 758–762, Aug. 2006.

[8] J. Wu, J. Long, and M. Liu, "Evolving RBF neural networks for rainfall prediction using hybrid particle swarm optimization and genetic algorithm," *Neurocomputing*, vol. 148, no. 2, pp. 136–142, Feb. 2015.

[9] Q.-H. Lin, F.-L. Yin, T.-M. Mei, and H. Liang, "A blind source separation based method for speech encryption," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 53, no. 6, pp. 1320–1328, Jun. 2006.

[10] D. Peng and Y. Xiang, "Underdetermined blind source separation based on relaxed sparsity condition of sources," *IEEE Trans. Signal Process.*, vol. 57, no. 2, pp. 809–814, Feb. 2009.

[11] Y. Deville, M. Benali, and F. Abrard, "Differential source separation for underdetermined instantaneous or convolutive mixtures: Concept and algorithms," *Signal Process.*, vol. 84, no. 10, pp. 1759–1776, Oct. 2004.

[12] J. Thomas, Y. Deville, and S. Hosseini, "Differential fast fixed-point algorithms for underdetermined instantaneous and convolutive partial blind source separation," *IEEE Trans. Signal Process.*, vol. 55, no. 7, pp. 3717–3729, Jul. 2007.

[13] Y. Xiang, S. K. Ng, and V. K. Nguyen, "Blind separation of mutually correlated sources using precoders," *IEEE Trans. Neural Netw.*, vol. 21, no. 1, pp. 82–90, Jan. 2010.

[14] Y. Li, J. Wang, and J. M. Zurada, "Blind extraction of singularly mixed source signals," *IEEE Trans. Neural Netw.*, vol. 11, no. 6, pp. 1413–1422, Nov. 2000.

[15] Y. Li, S.-I. Amari, A. Cichocki, D. W. C. Ho, and S. Xie, "Underdetermined blind source separation based on sparse representation," *IEEE Trans. Signal Process.*, vol. 54, no. 2, pp. 423–437, Feb. 2006.

[16] R. Gribonval and S. Lesage, "A survey of sparse component analysis for blind source separation: Principles, perspectives, and new challenges," in *Proc. ESANN*, vol. 1. 2006, pp. 323–330.

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

[18] D. Luengo, I. Santamaria, J. Ibanez, L. Vielva, and C. Pantaleon, "A fast blind SIMO channel identification algorithm for sparse sources," *IEEE Signal Process. Lett.*, vol. 10, no. 5, pp. 148–151, May 2003.

[19] P. Georgiev, F. Theis, and A. Cichocki, "Sparse component analysis and blind source separation of underdetermined mixtures," *IEEE Trans. Neural Netw.*, vol. 16, no. 4, pp. 992–996, Jul. 2005.

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

[21] A. Jourjine, S. Rickard, and O. Yilmaz, "Blind separation of disjoint orthogonal signals: Demixing N sources from 2 mixtures," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, vol. 5. Jun. 2000, pp. 2985–2988.

[22] D. Peng, Y. Xiang, and H. Trinh, "A new blind method for separating $M+1$ sources from M mixtures," *Comput. Math. Appl.*, vol. 60, no. 7, pp. 1829–1839, Oct. 2010.

[23] D. Peng and Y. Xiang, "Underdetermined blind separation of non-sparse sources using spatial time-frequency distributions," *Digit. Signal Process.*, vol. 20, no. 2, pp. 581–596, Mar. 2010.

[24] N. Linh-Trung, A. Belouchrani, K. Abed-Meraim, and B. Boashash, "Separating more sources than sensors using time-frequency distributions," *EURASIP J. Appl. Signal Process.*, vol. 2005, no. 17, pp. 2828–2847, Dec. 2005.

[25] S. Xie, L. Yang, J.-M. Yang, G. Zhou, and Y. Xiang, "Time-frequency approach to underdetermined blind source separation," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 23, no. 2, pp. 306–316, Feb. 2012.

[26] A. Aissa-El-Bey, N. Linh-Trung, K. Abed-Meraim, A. Belouchrani, and Y. Grenier, "Underdetermined blind separation of nondisjoint sources in the time-frequency domain," *IEEE Trans. Signal Process.*, vol. 55, no. 3, pp. 897–907, Mar. 2007.

[27] P. S. Bradley and O. L. Mangasarian, "k-plane clustering," *J. Global Optim.*, vol. 16, no. 1, pp. 23–32, 2000.

[28] R. Vidal, "Subspace clustering," *IEEE Signal Process. Mag.*, vol. 28, no. 2, pp. 52–68, Mar. 2011.

[29] F. Abrard, Y. Deville, and P. White, "From blind source separation to blind source cancellation in the underdetermined case: A new approach based on time-frequency analysis," in *Proc. 3rd Int. Conf. Independ. Compon. Anal. Signal Separat. (ICA)*, San Diego, CA, USA, 2001, pp. 734–739.

[30] F. Abrard and Y. Deville, "Blind separation of dependent sources using the 'time-frequency ratio of mixtures' approach," in *Proc. 7th Int. Symp. Signal Process. Appl. (ISSPA)*, Paris, France, Jul. 2003, pp. 1–4.

[31] F. Abrard and Y. Deville, "A time–frequency blind signal separation method applicable to underdetermined mixtures of dependent sources," *Signal Process.*, vol. 85, no. 7, pp. 1389–1403, Jul. 2005.

[32] A. Javanmard, P. Pad, M. Babaie-Zadeh, and C. Jutten, "Estimating the mixing matrix in underdetermined Sparse Component Analysis (SCA) using consecutive independent component analysis (ICA)," in *Proc. 16th Eur. Signal Process. Conf. (EUSIPCO)*, Lausanne, Switzerland, Sep. 2008, pp. 1–5.

[33] D. L. Donoho and M. Elad, "Optimally sparse representation in general (nonorthogonal) dictionaries via ℓ^1 minimization," *Proc. Nat. Acad. Sci. USA*, vol. 100, no. 5, pp. 2197–2202, 2003.

[34] M. S. Karoui, Y. Deville, S. Hosseini, and A. Ouamri, "Blind spatial unmixing of multispectral images: New methods combining sparse component analysis, clustering and non-negativity constraints," *Pattern Recognit.*, vol. 45, no. 12, pp. 4263–4278, 2012.

[35] J.-F. Cardoso and A. Souloumiac, "Blind beamforming for non-Gaussian signals," *IEE Proc.-F Radar Signal Process.*, vol. 140, no. 6, pp. 362–370, Dec. 1993.

[36] L. Sorber, M. Van Barel, and L. de Lathauwer. (Feb. 2013). *Tensorlab V1.0*. [Online]. Available: http://esat.kuleuven.be/sista/tensorlab/

PENG CHEN received the B.Sc. degree in computer science from the University of Electronic Science and Technology of China, Chengdu, China, in 2001, the M.Sc. degree in computer science from Peking University, Beijing, China, in 2004. He is currently pursuing the Ph.D. degree with the Machine Intelligence Laboratory, College of Computer Science, Sichuan University, Chengdu, China. His research interests include blind signal processing and neural networks.

DEZHONG PENG (M'09) received the B.Sc. degree in applied mathematics, the M.Sc. and Ph.D. degrees in computer software and theory from the University of Electronic Science and Technology of China, Chengdu, China, in 1998, 2001, and 2006, respectively. From 2001 to 2007, he was with the University of Electronic Science and Technology of China as an Assistant Lecturer and a Lecturer. He was a Postdoctoral Research Fellow with the School of Engineering, Deakin University, Australia from 2007 to 2009. He is currently a Professor with the Machine Intelligence Laboratory, College of Computer Science, Sichuan University, Chengdu, China. His research interests include blind signal processing and neural networks.

LIANGLI ZHEN received the B.Eng. degree in software engineering from the Southwest University of Science and Technology and the M.Eng. degree in computer science from the Sichuan University, China, in 2011 and 2014, respectively. He is currently pursuing the Ph.D. degree with the Machine Intelligence Laboratory, College of Computer Science, Sichuan University, Chengdu, China and also a visiting Ph.D. student with the Centre of Excellence for Research in Computational Intelligence and Applications, School of Computer Science, University of Birmingham, U.K. His research interests include machine learning and evolutionary computation.

YIFAN LUO received the B.Sc. degree in computer science and technology from the University of Electronic Science and Technology of China, Chengdu, China, in 2005, the M.Sc. in educational technology from the Sichuan Normal University, Chengdu, China, in 2009, respectively. He is currently pursuing the Ph.D. degree with the Machine Intelligence Laboratory, College of Computer Science, Sichuan University, Chengdu, China. Since 2009, he was an Assistant Lecturer with Sichuan Post and Telecommunication College and a Lecturer. His research interest is digital watermarking.

YONG XIANG (SM'12) received the Ph.D. degree in electrical and electronic engineering from The University of Melbourne, Australia. He is a Professor and the Director with the Artificial Intelligence and Data Analytics Research Cluster, School of Information Technology, Deakin University, Australia. His research interests include information security and privacy, multimedia (speech, image, and video) processing, wireless sensor networks and IoT, and biomedical signal processing. He has published two monographs, over 90 refereed journal articles, and numerous conference papers in these areas. He is an Associate Editor of the IEEE SIGNAL PROCESSING LETTERS and the IEEE ACCESS. He has served as Program Chair, TPC Chair, Symposium Chair, and Session Chair for a number of international conferences.

● ● ●