Received April 26, 2019, accepted May 28, 2019, date of publication June 10, 2019, date of current version July 15, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2921919

Underdetermined Blind Source Separation With Multi-Subspace for Nonlinear Representation

LU WANG[1], (Student Member, IEEE), AND TOMOAKI OHTSUKI[2], (Senior Member, IEEE)
[1]Graduate School of Science and Technology, Keio University, Yokohama 223-8522, Japan
[2]Department of Information and Computer Science, Keio University, Yokohama 223-8522, Japan

Corresponding author: Lu Wang (wanglu@ohtsuki.ics.keio.ac.jp)

This work was supported by the Keio Leading-Edge Laboratory of Science and Technology, KEIO-KLL-000045 Japan.

ABSTRACT Blind source separation (BSS) is a technique to recognize the multiple talkers from the multiple observations received by some sensors without any prior knowledge information. The problem is that the mixing is always complex, such as the case where sources are mixed with some direction angles, or where the number of sensors is less than that of sources. In this paper, we propose a multi-subspace representation based BSS approach that allows the mixing process to be nonlinear and underdetermined. The approach relies on a multi-subspace structure and sparse representation in the time-frequency (TF) domain. By parameterizing such subspaces, we can map the observed signals in the feature space with the coefficient matrix from the parameter space. We then exploit the linear mixture in the feature space that corresponds to the nonlinear mixture in the input space. Once such subspaces are built, the coefficient matrix can be constructed by solving an optimization problem on the coding coefficient vector. Relying on the TF representation, the target matrix can be constructed in a sparse mixture of TF vectors with the fewer computational cost. The experiments are designed on the observations that are generated from an underdetermined mixture, and that is collected with some direction angles in a virtual room environment. The proposed approach exhibits higher separation accuracy.

INDEX TERMS Underdetermined BSS, multi-subspace representation, nonlinear mixture, sparse coding, time-frequency representation.

I. INTRODUCTION

Recognizing multiple talkers from the multiple observations (or mixtures) received by a set of sensors is the task of source separation. The problem is referred to as ''blind'' source separation when the procedure has access only to the observations without any prior knowledge information for the mixing system. In general, most BSS algorithms assume that the number of sources is less than that of sensors, denoted as overdetermined BSS. However, in practice, this assumption is difficult to be satisfied since the number of sources is unknown.

Various attempts [1]–[3] on underdetermined BSS (UBSS) have been proposed that consider the scenario, where the number of sensors is less than that of sources. Since the mixing matrix is irreversible in this case, the recovered sources also need to be estimated even though the mixing matrix has been known. To solve this problem, a well-known framework has been proposed by exploiting the sparseness of the sources in the representation domain, such as wavelet packet transform [4] or short-time Fourier transform (STFT) [5]. For instance, the degenerate unmixing estimation technique (DUET) was proposed in [6]. The approach exploits the ratio of TF transforms of the observed signals to recover the source signals. Yilmaz and Rickard [7] assumed that the sources are disjoint in the TF domain. These methods work on the assumption that there exists at most one active source at any point in the TF domain. This implies that the separation performance will degrade as the number of the TF disjoints points being increased. To relax this constraint, [8], [9] proposed a scenario that allows the sources to be non-disjoint in the TF domain, however, the number of the sources that coexist at any TF point is less than that of the mixtures [8].

In the above methods, the mixing process is considered to be linear only. In fact, however, the assumption is restrictive and easy to be violated in the real-world applications [10], such as communication [11], [12], speech or audio processing [13], and biomedical engineering [14].

The associate editor coordinating the review of this manuscript and approving it for publication was Alma Y. Alanis.

The problem for the nonlinear BSS is intractable solely based on the assumption that the sources are statistically independent. e.x., if x and y are two independent random variables, then $f(x)$ and $g(y)$ are also independent for any f and g [15]. Therefore, the solutions are highly non-unique without any further constraints for the space of nonlinear mixing function [16].

Efforts on exploiting such further constraints in the nonlinear domain have involved, such as extracting unknown nonlinearities upon unknown parameters [17], approximating a nonlinear function whose inverse function can be constrained well on the estimator of a priori neural network [11], [18]. Another popular approach consists in using kernel so as to implicitly map the data via kernel trick. The main advantage of this approach is that the estimation of the parameters in the model is actually independent of the number of channels. Formally, the data are mapped into $\mathcal{H}$ using $\phi : \mathcal{X} \rightarrow \mathcal{H}, \mathbf{x} \rightarrow \phi(\mathbf{x})$ so as to extract the nonlinearity. To avoid working on the high-dimensional space $\mathcal{H}$, one tries in the feature space in which the dot product can be calculated by $k(\mathbf{x}, \mathbf{x}') = \langle \phi(\mathbf{x}), \phi(\mathbf{x}') \rangle$, which is called as kernel trick.

Typically, Harmeling *et al.* [19], and Martinez and Bray [20] exploit the temporal information of sources for separation, and do not enforce mutual independence of outputs. This method produces successful results in many experiments. However, a problem is that the cost of storing and evaluating the model is proportional to the number of data points [21]. Moreover, this method may fail if some sources lack specific time structures. [22], [23] provides a good approximation of the value attained by the nonlinear mixing. Relying on such spaces spanned by a set of vanishing polynomial, the data implicitly mapped into high-dimensional space, and the effective subspace is extracted. It allowed us to solve a nonlinear problem linearly. But, unfortunately, the approach can not be used for the underdetermined case.

In this paper, we propose a multi-subspace representation based separation approach that tackles the scenario of the nonlinear and underdetermined mixture. The separation system is constructed using the kernel methods with a multi-subspace structure. To obtain a set of basis so as to the spanned subspace could be orthonormal in the theoretical support, we propose to use the geometric vertices of data. Then we solve a linear problem by exploiting the technique of sparse coding. The coefficient matrix is adjusted by minimizing the loss function.

We first consider a model related to the input space $\mathbf{x} \in \mathbb{R}^N$ by a kernel mapping with multi-subspace structure. The effective number of basis denoted by k, provides the smallest construction error in the nonlinear approximation. One of the keys in that algorithm is to find a set of orthogonal basis to study the parameterized signals in multiple feature spaces. Some techniques [19], [24], [25] can help that are roughly analogous. Either random sampling or k-means clustering is considered to obtain some vectors, which is expected to be independent. However, the method may not be appropriate for mixture data. We attempt to use the geometric vertices of the convex hull as the basis, which parameterizes the multi-subspace that contains the reduced vectors in the feature space. Relying on a set of an orthonormal basis, the spanned subspaces can represent the nonlinearity of mixing function in the minimum number.

Another contribution is to derive the coefficient matrix by solving the loss function on the coding coefficient vector. Once such subspaces are built, by allowing multiple sources to be presented at any point in the TF domain, we can figure out the target matrix in a sparse mixture TF vectors with less computational cost. Finally, using this coefficient matrix, the original sources in underdetermined scenarios can be estimated.

The remainder of this paper is organized as follows. Section II reviews the consents of convex geometry, Kernel theorem first. Then, the nonlinear mixture model is introduced for further study. Section III introduces some conditions necessary for the separation of nonstationary sources in the TF domain. Section IV describes our proposed separation approach that relies on multi-subspaces representation and sparse representation in the TF domain. Section V shows the experimental settings and results. Conclusions are reported in Section VI.

II. NOTATIONS AND SYSTEM MODEL

The following notations will be used in the ensuing presentation.

$\mathbf{A}, \mathbf{A}^\top, \mathbf{A}^\dagger$	Matrix, its transposed matrix and pseudo-inverse matrix.
$\mathbb{R}, \mathbb{R}^N, \mathbb{R}^{N\times M}$	Set of real numbers, set of N vectors, and set of $N \times M$ matrices.
$\mathbb{R}_+, \mathbb{R}_+^N, \mathbb{R}_+^{N\times M}$	Set of non-negative real numbers, set of non-negative N vectors, and set of non-negative $N \times M$ matrices.
$\mathbf{1}_N$	Vector one of N elements.
$\mathbf{s}, \mathcal{D}_\mathbf{s}$	The vector of source signals, and source signals in the TF domain.
$\mathbf{x}, \mathcal{D}_\mathbf{x}$	The vector of observed signals, and observed signals in the TF domain.

In the following, a brief review of some concepts on convex geometry and Kernel method will be given for ease of later use.

A. CONVEX GEOMETRY

The Definition 1 of convex hull [26] for a set of vectors $\{\mathbf{x}(1), \cdots, \mathbf{x}(T)\}$ will be given in the following.

Definition 1: Given a set of vectors $\mathcal{X} = \{\mathbf{x}(1), \cdots, \mathbf{x}(T)\}$. The convex hull of the finite nonempty set $\mathcal{X} \subseteq \mathbb{R}^N$ gives the form

$$\operatorname{conv}\{\mathbf{x}(1), \cdots, \mathbf{x}(T)\} = \left\{ \sum_{i=1}^{T} \lambda_i \mathbf{x}(i) \middle| \boldsymbol{\lambda} \in \mathbb{R}_+^T, \mathbf{1}_T^\top \boldsymbol{\lambda} = 1 \right\},$$

where $\boldsymbol{\lambda} = [\lambda_1, \cdots, \lambda_T]^\top$ is any non-negative vector. □

In the above equation, conv$\{\mathbf{x}(1), \cdots, \mathbf{x}(T)\}$ is called as a $(T-1)$-dimensional simplex with T vertices $\{\mathbf{x}(1), \cdots, \mathbf{x}(T)\}$ if and only if $\{\mathbf{x}(1), \cdots, \mathbf{x}(T)\}$ is affinely independent, or equivalently. Furthermore, if $\{\mathbf{x}(1) - \mathbf{x}(T), \cdots, \mathbf{x}(T-1) - \mathbf{x}(T)\}$ is linearly independent that is called a simplest simplex in $\mathbb{R}^N$ [27]. As see in the Fig. 1, a triangle is a 2-dimensional simplest simplex in $\mathbb{R}^2$, and a tetrahedron is a 3-dimensional simplest simplex in $\mathbb{R}^3$.

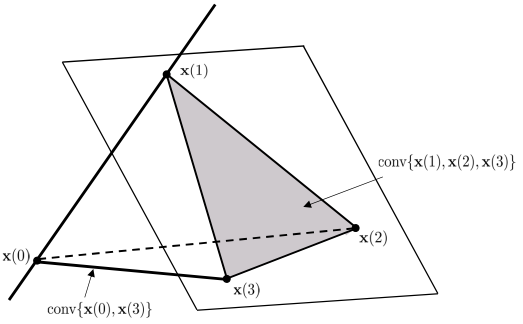

FIGURE 1. A graphical illustration for the convex geometry concepts. The line segment connecting x(0) and x(3) is the convex hull of {x(0), x(3)}, which is denoted by conv{x(0), x(3)}. The shaded triangle is the convex hull of {x(1), x(2), x(3)}, i.e., conv{x(1), x(2), x(3)}.

According to the N-FINDR criterion [28], the approach finds the endmembers' convex hull that in fact of extracting the data-enclosing simplex with the maximum volume [29]. That can be given by solving the maximization problem

$$\max_{\mathbf{p}(i)\in\mathbb{R}^{M-1},\forall i} \mathcal{V}(\mathbf{p}(1), \cdots, \mathbf{p}(k)) \quad \text{s.t. } \mathbf{x}(t) \in \text{conv}\{\mathbf{p}(1), \cdots, \mathbf{p}(k)\}, \quad \forall t$$

where $\mathcal{V}(\cdot)$ denotes the volume of the simplex conv$\{\mathbf{p}(1), \cdots, \mathbf{p}(k)\} \subseteq \mathbb{R}^{M-1}$.

The above theory is introduced for the theoretical support in our further work, where the geometric vertices can establish a set of orthogonal basis so that the spanned multiple subspaces can represent the nonlinearity in the minimum number.

B. NONLINEAR MIXTURE MODEL

Consider a nonlinear, instantaneous and invertible mixing system with M inputs and N outputs

$$\mathbf{x}(t) = \mathcal{F}(\mathbf{s}(t)), \tag{1}$$

for $t = 1, 2, \cdots, T$, where $\mathbf{s}(t) = [s_1(t), \cdots, s_M(t)]^\top$ is the original sources of M statistically independent vectors. The superscript $[\cdot]^\top$ denotes the transpose operator. $s_i(t)$ denotes the original source of the i-th signal at t time index. The mixing function $\mathcal{F}$ transform the $\mathbf{s}(t)$ from $\mathbb{R}^M$ to $\mathbb{R}^N$, i.e., the observations $\mathbf{x}(t) = [x_1(t), \cdots, x_N(t)]^\top$ are N-dimensional mixture vectors.

The general idea of performing is to design a separation function $\mathcal{G} : \mathbb{R}^N \to \mathbb{R}^M$ such that

$$\hat{\mathbf{s}}(t) = \mathcal{G}(\mathbf{x}(t)), \tag{2}$$

where the recovered sources $\hat{\mathbf{s}}$ are statistically independent. One has been given in [30], where the nonlinear mixtures of independent variables are still independent. However, the statistical independence of estimated sources is no longer a sufficient constraint for demixing function, without additional prior knowledge on the mixing process [16]. To form a mapping function with multi-subspace structure, we consider the Kernel theorem and its feature space.

C. KERNEL AND FEATURE SPACE

The key point is how to generate a mapping function that can achieve the approximation of the inverse operator of (1). In [19], the kernelization method was introduced by mapping the data $\mathbf{x}(t)$ implicitly into the kernel feature space $\mathcal{H}$ with the kernel function $\mathcal{K} : \mathbb{R}^N \times \mathbb{R}^N \to \mathbb{R}$. The basic definitions are introduced at first.

Definition 2: Let $\mathcal{X}$ be a nonempty set. The symmetric function $\mathcal{K} : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ is called as a positive definite kernel, if

$$\sum_{i,j=1}^{N} c_i c_j \mathcal{K}(\mathbf{x}(i), \quad \mathbf{x}(j)) \geq 0, \tag{3}$$

holds for any $\mathbf{x}(i) \in \mathcal{X}$ and $c_1, \cdots, c_N \in \mathbb{R}$. □

One can easily deduce from Definition 2 that the positive definite kernel transforms data into kernel feature space, which can be simply calculated by matrices of kernel built on the sample of points as

$$\langle \boldsymbol{\phi}(\mathbf{x}(i)), \boldsymbol{\phi}(\mathbf{x}(j)) \rangle = \mathcal{K}(\mathbf{x}(i), \mathbf{x}(j)), \tag{4}$$

where $i, j = 1, \cdots, T$ and $\langle \cdot, \cdot \rangle$ is the inner product. $\boldsymbol{\phi}(\mathbf{x})$ is the Hilbert mapping function. Using the kernel trick, the inner product of two feature mappings in the Hilbert space can be computed by a kernel function in the original space. The computational complexity can be controlled within a linear range.

This would first define a direction $\mathbf{W} \in \mathcal{H}$ that enables us to parameterize the data by

$$\mathbf{W} = \boldsymbol{\Phi}_{\mathbf{x}} \boldsymbol{\alpha} = \sum_{j=1}^{T} \alpha_j \boldsymbol{\phi}(\mathbf{x}(j)) \in \mathcal{H}, \tag{5}$$

where $\boldsymbol{\alpha} = [\alpha_1, \cdots, \alpha_T]^\top$ is a parameter vector. $\boldsymbol{\Phi}_{\mathbf{x}}$ is the matrix with the column vectors $[\boldsymbol{\phi}(\mathbf{x}(1)), \cdots, \boldsymbol{\phi}(\mathbf{x}(T))]^\top$. Using the kernel trick of (4), the demixing process in the feature space is given by

$$\begin{aligned} \hat{\mathbf{s}}(t) &= \mathbf{W}^\top \boldsymbol{\Phi}(\mathbf{x}(t)) = \boldsymbol{\alpha}^\top \boldsymbol{\Phi}_{\mathbf{x}}^\top \boldsymbol{\phi}(\mathbf{x}(t)) \\ &= \sum_{j=1}^{T} \alpha_j \mathcal{K}(\mathbf{x}(j), \mathbf{x}(t)). \end{aligned} \tag{6}$$

The main advantage of Kernel mapping is that the number of parameters to estimate in the model is actually independent of the number of channels. However, without extra constraints, generating a unique mapping function is intractable.

This paper proposes a multi-subspace representation based on Kernel spaces to tackle the ill-posed with a few assumptions. The k multiple subspaces produce k outputs, and we propose the way to select n outputs as the estimator of the original sources.

III. LINEAR TF-UBSS APPROACH

We first review the TF domain based underdetermined BSS (UBSS) method that was presented by [8] and later proposes a multi-layer representation based nonlinear TF-UBSS algorithm. The discrete-time short-time Fourier transform (STFT) is given by

$$\mathcal{D}_{s_i}(\tau,\omega)=\sum_{t=-\infty}^{\infty}s_i(t)h(t-\tau)e^{-j\omega t}, \tag{7}$$

at frame τ and frequency bin ω, where $h(t)$ is a window function. Using STFT of (7), the linear BSS can be transformed into the TF domain

$$\mathcal{D}_{\mathbf{x}}(t,\omega)=\mathbf{A}\mathcal{D}_{\mathbf{s}}(t,\omega), \tag{8}$$

where $\mathcal{D}_{\mathbf{x}}(t,\omega)=[\mathcal{D}_{x_1}(t,\omega),\cdots,\mathcal{D}_{x_N}(t,\omega)]^\top$ is the mixture signals in the TF domain and $\mathcal{D}_{\mathbf{s}}(t,\omega)=[\mathcal{D}_{s_1}(t,\omega),\cdots,\mathcal{D}_{s_M}(t,\omega)]^\top$ is the STFT vector of the source signals. $\mathcal{D}_{s_i}(t,\omega)$ is the i-th source signal in the ω-th frequency bin at t time index.

Assumption 1: For each source signal $\mathbf{s}_i$, its STFT transformation is denoted as $\mathcal{D}_{\mathbf{s}_i}$ in the TF domain. There are some TF points, where only $\mathbf{s}_i$ is dominant, i.e., $|\mathcal{D}_{\mathbf{s}_i}(t,\omega)|\gg|\mathcal{D}_{\mathbf{s}_j}(t,\omega)|$ for $\forall j\neq i$.

The assumption implies that all sources are disjoint in the TF domain, i.e., there is only one source that is active. Then, (8) can be rewritten as

$$\mathcal{D}_{\mathbf{x}}(t_a,\omega_a)=\mathbf{a}_i\mathcal{D}_{\mathbf{s}_i}(t_a,\omega_a), \tag{9}$$

where the subscript a indicates any one of the sources is active in the TF domain.

The noise thresholding procedure proposed by [7] is used to keep those points having sufficient energy, which is referred to as auto-source points. The procedure is performed for each time-slice of the TF representation, by applying a criterion for all the frequency points belonging to this time-slice

$$\text{if}\ \frac{\|\mathcal{D}_{\mathbf{x}}(t_a,\omega_a)\|}{\max_\omega\{\|\mathcal{D}_{\mathbf{x}}(t_a,\omega)\|\}}>\epsilon,\quad\text{then keep }(t_a,\omega_a), \tag{10}$$

where ϵ is a small threshold, e.x., the threshold $\epsilon=0.05$ is given in [8]. Then, the set of all selected points Ω is expressed by $\Omega=\bigcup_{i=1}^n\Omega_i$, where Ω_i is the TF support of the source $s_i(t)$.

To estimate the mixing vectors $\mathbf{a}_i$, the clustering algorithm is performed on the assumption in [8] that the highest densities occur around the vectors $\mathbf{a}_i$. Thus, the average values over the samples of each cluster are defined as the mixing vectors

$$\hat{\mathbf{a}}_i=\frac{1}{|C_i|}\sum_{(t,\omega)\in\Omega_i}\frac{\mathcal{D}_{\mathbf{x}}(t,\omega)}{\|\mathcal{D}_{\mathbf{x}}(t,\omega)\|}, \tag{11}$$

where $|C_i|$ is the number of vectors included in the same cluster.

Finally, each source in the TF domain can be estimated by

$$\hat{\mathcal{D}}_{s_i}(t,\omega)=\begin{cases}\hat{\mathbf{a}}_i^\dagger\mathcal{S}_{\mathbf{x}}(t,\omega), & \forall(t,\omega)\in\Omega_i,\\ 0, & \text{otherwise},\end{cases} \tag{12}$$

where the superscript $[\cdot]^\dagger$ denotes the pseudo-inverse operator. The source estimator $\hat{s}_i(t)$ is then obtained by transforming $\hat{\mathcal{D}}_{s_i}(t,\omega)$ into the time domain using the inverse STFT.

IV. MULTI-SUBSPACE REPRESENTATION BASED NONLINEAR TF-UBSS APPROACH

The TF-UBSS method relies on the assumption that the sources were mixed linearly, which has led to the recovered structure in (12). However, for the nonlinear blind source separation, the solutions are non-unique [16] without any extra constraints for the mixing process. In this paper, we propose a multi-subspace representation to construct the nonlinear variants by mapping the data implicitly in some kernel feature spaces. If one of the subspaces can match the nonlinearity of the mixing functions, the nonlinear problem can be broken down into the version of the linear case.

A. CHOOSING VECTORS FOR BASIS

To extract a vector that formed a matrix with full column rank, we use the N-FINDR algorithm, which was originally developed by Winter in [28]. The approach finds a set of vertices in fact of extracting a vector of data space that defined the largest volume.

Definition 3: Let $\mathcal{X}=\{\mathbf{x}(i)\}_{i=1}^T$ be a set of sample vectors. The convex hull of the finite nonempty set $\mathcal{X}\subseteq\mathbb{R}^d$ gives the form

$$\text{conv}(\{\mathbf{x}(1),\cdots,\mathbf{x}(T)\})=\left\{\sum_{i=1}^T\lambda_i\mathbf{x}(i)\Big|\lambda_i\geq0,\sum_i\lambda_i=1\right\}, \tag{13}$$

Proposition 1: Let $\{\mathbf{p}(1),\cdots,\mathbf{p}(k)\}$ be a subset of vectors in the convex hull $\mathcal{X}=\{\mathbf{x}(i)\}_{i=1}^T$. For $k\ll T$, if the vectors $\mathbf{p}(1),\cdots,\mathbf{p}(k)$ are the vertices of $\mathcal{X}$, then we have

$$\text{conv}(\{\mathbf{x}(1),\cdots,\mathbf{x}(T)\})\subseteq\text{conv}(\{\mathbf{p}(1),\cdots,\mathbf{p}(k)\}). \tag{14}$$

Proof: Without loss of generality, $\mathbf{x}(1),\cdots,\mathbf{x}(k)$ are the vertices of $\mathcal{P}:=\text{conv}(\{\mathbf{x}(1),\cdots,\mathbf{x}(T)\})$, which are expressed as $\mathbf{p}(1),\cdots,\mathbf{p}(k)$. For any $i>k$, if $\mathbf{x}(i)$ is not a vertex of $\mathcal{P}$, then $\mathbf{x}(i)$ can be expressed by a linear combination $\mathbf{x}(i)=\sum_{j=1}^k\lambda_j\mathbf{p}(j)$. Thus, for any sample $\mathbf{x}\in\mathcal{P}$, we have

$$\begin{aligned}\mathbf{x}&=\sum_{i=1}^T\mu_i\mathbf{x}(i)=\sum_{i=1}^k\mu_i\mathbf{p}(i)+\sum_{i=k+1}^T\mu_i\mathbf{x}_i\\&=\sum_{i=1}^k\mu_i\mathbf{p}(i)+\sum_{i=k+1}^T\mu_i\sum_{j=1}^k\lambda_j\mathbf{p}(j)\\&=\sum_{i=1}^k\left(\mu_i+\lambda_i\sum_{j=k+1}^T\mu_j\right)\mathbf{p}(i).\end{aligned} \tag{15}$$

Since $\sum_{i=1}^{k}(\mu_i + \lambda_i \sum_{j=k+1}^{T} \mu_j) = 1$, we conclude that $\mathbf{x} \in \text{conv}(\{\mathbf{p}(1), \cdots, \mathbf{p}(T-1)\}) \subseteq \text{conv}(\{\mathbf{p}(1), \cdots, \mathbf{p}(k)\})$. ■

Proposition 1 implies that the volume simplex formed by the vertices is larger than or equal to any other volume defined by any other combination of elements. Thus, the vertices can be extracted in fact of finding a vector of data space that formed the maximum volume. The approach can be briefly described in the following implementation.

For a vertex simplex composed of k vectors $\mathbf{p}(1), \mathbf{p}(2), \cdots, \mathbf{p}(k)$, its volume $\mathcal{V}(\mathbf{P}) = \mathcal{V}(\mathbf{p}(1), \cdots, \mathbf{p}(k))$ is defined by

$$\mathcal{V}(\mathbf{P}) = \frac{\left|\det\begin{bmatrix} 1 & \cdots & 1 \\ \mathbf{p}(1) & \cdots & \mathbf{p}(k) \end{bmatrix}\right|}{(k-1)!}. \tag{16}$$

Find a set of k vectors in the data, denoted by $\mathbf{P}^* = [\mathbf{p}^*(1), \mathbf{p}^*(2), \cdots, \mathbf{p}^*(k)]$, that forms a k-vertex simplex to yield the maximum value of (16), which is given by

$$\{\mathbf{p}^*(i_1), \cdots, \mathbf{p}^*(i_k)\} = \arg \max_{\mathbf{p}(i_1), \cdots, \mathbf{p}(i_k)} \mathcal{V}(\mathbf{P}). \tag{17}$$

Thus, the desired set of independent vectors $\{\mathbf{p}^*(i_1), \mathbf{p}^*(i_2), \cdots, \mathbf{p}^*(i_k)\}$ are found. Assume that the dimension of vector $\mathbf{p}^*$ is larger that the number of vector k, then the columns of the matrix being linearly independent. For further work, a set of orthonormal subspaces produced by these k vectors can represent the nonlinearity or distortion caused by the mixing functions using the reduced data.

B. CONSTRUCTING A MULTI-SUBSPACE REPRESENTATION

Given the observation data $\mathbf{x}(t) \in \mathbb{R}^N$, for all $t = 1, \cdots, T$ that are assumed to be generated by the nonlinear mixture functions. To make the nonlinear problem linearly separable, the idea is to fulfill a certain condition that induces a mapping $\Phi : \mathbb{R}^N \to \mathcal{H}$ in the feature space. Therefore, we attempt to find some mapping functions, which are used to capture the varieties of nonlinearity or distortion.

To describe the nonlinearity efficiently in a feature space, we use a subset from $\{\mathbf{x}(t)\}_{t=1}^{T} \in \mathbb{R}^N$, denoted as $\mathbf{p}(1), \cdots, \mathbf{p}(k) \in \mathbb{R}^N$ to generate a set of basis in $\mathcal{H}$. Since the data points belonging to the subset is expected to be mutually independent in the feature space, we use the k center points of clusters to form the subset $\{\mathbf{p}(i)\}_{i=1}^{k}$. Thus, we can define an orthonormal basis by using the empirical kernel map

$$\mathbf{\Xi} := \mathbf{\Phi_p} \langle \mathbf{\Phi_p}, \mathbf{\Phi_p} \rangle^{-\frac{1}{2}}, \tag{18}$$

where $\mathbf{\Phi_p} = [\Phi(\mathbf{p}_1), \cdots, \Phi(\mathbf{p}_k)]$ is the mapping of data points in the feature space.

By defining the basis that allows us to parameterize such subspace, the observed signals are mapped in the feature space with the coefficient matrix from a parameter space.

$$\begin{aligned} &\mathbf{\Psi}(\mathbf{x}(t)) \\ &= \mathbf{\Xi}^\top \mathbf{\Phi}(\mathbf{x}(t)) = \langle \mathbf{\Phi_p}, \mathbf{\Phi_p} \rangle^{-\frac{1}{2}} \langle \mathbf{\Phi_p}, \mathbf{\Phi}(\mathbf{x}(t)) \rangle \\ &= \begin{bmatrix} \mathcal{K}(\mathbf{p}(1), \mathbf{p}(1)) & \cdots & \mathcal{K}(\mathbf{p}(1), \mathbf{p}(k)) \\ \vdots & & \vdots \\ \mathcal{K}(\mathbf{p}(k), \mathbf{p}(1)) & \cdots & \mathcal{K}(\mathbf{p}(k), \mathbf{p}(k)) \end{bmatrix}^{\frac{1}{2}} \begin{bmatrix} \mathcal{K}(\mathbf{p}(1), \mathbf{x}[t]) \\ \vdots \\ \mathcal{K}(\mathbf{p}(k), \mathbf{x}[t]) \end{bmatrix} \end{aligned} \tag{19}$$

where $\mathcal{K}(\mathbf{p}(i), \mathbf{p}(j))_{i,j}^{-\frac{1}{2}}$ is an invertible real valued matrix. Due to the $\mathbf{\Phi_p}$ constructed by a subset, the computational complexity of the projection function in (19) is reduced to $\mathcal{O}(k^2N) + \mathcal{O}(kNT) + \mathcal{O}(k^2T)$ from original $\mathcal{O}(T^2N) + \mathcal{O}(NT^2) + \mathcal{O}(T^3)$, where $T \gg k$.

Thus, the demixing process can be defined in the feature space as

$$\hat{\mathbf{s}}(t) = \mathbf{W}^\dagger \mathbf{\Psi}(\mathbf{x}(t)). \tag{20}$$

The above equation implies that the nonlinear problem can be linearly separable in the feature space.

C. COEFFICIENT MATRIX IDENTIFICATION

Relying on the linear relation of (20), we have the corresponding representation by using STFT,

$$\mathcal{D}_{\mathbf{\Psi}}(t, \omega) = \tilde{\mathbf{W}} \hat{\mathcal{D}}_{\mathbf{s}_i}(t, \omega). \tag{21}$$

Based on Assumptions 1, we know that there exists only one estimated source $\hat{\mathbf{s}}_i$ being active on the TF point (t, ω). Then, we have

$$\mathcal{D}_{\mathbf{\Psi}}(t, \omega) = \hat{\mathcal{D}}_{\mathbf{s}_i}(t, \omega) \tilde{\mathbf{W}}_i, \tag{22}$$

where the TF feature matrix $\mathcal{D}_{\mathbf{\Psi}}(t, \omega)$ can be represented by the i-th column vector $\tilde{\mathbf{W}}_i$ up to a multiplicative coefficient $\mathcal{D}_{\hat{\mathbf{s}}_i}(t, \omega)$. This implies that the target matrix $\tilde{\mathbf{W}}_i$ can be a linear combination of a few numbers of sample points from the matrix $\mathcal{D}_{\mathbf{\Psi}}(t, \omega)$ with the coefficient $\hat{\mathcal{D}}_{\mathbf{s}_i}(t, \omega)$.

Thus, estimating a column vector of the coefficient matrix $\tilde{\mathbf{W}}_i$ can be achieved by finding the solution of a sparse representation $\mathcal{D}_{\mathbf{\Psi}}(t, \omega)$ with low-dimensional subspace. To remove the effect of noise, we use the criterion for all the frequency points belonging to this time-slice

$$\text{if } \frac{\|\mathcal{D}_{\mathbf{\Psi}}(t_p, \omega_k)\|}{\max_\omega \{\|\mathcal{D}_{\mathbf{\Psi}}(t_p, \omega)\|\}} > \epsilon, \quad \text{then keep } (t_p, \omega_k), \tag{23}$$

where ϵ is a small threshold, e.x., the threshold $\epsilon = 0.05$ is given in [8].

We next formulate the problem of (22) by using a sparse direction for TF representation of the mixture TF matrix $\mathcal{D}_{\mathbf{\Psi}}(t, \omega)$. Let $\boldsymbol{\pi}_1, \boldsymbol{\pi}_2, \cdots, \boldsymbol{\pi}_L$ be the reshaped vector of all the mixture TF matrix $\mathcal{D}_{\mathbf{\Psi}}$, and L is the number of TF points (t, ω). We can define a one row vector $\mathcal{D}_{\mathbf{\Pi}} \triangleq [\boldsymbol{\pi}_1, \cdots, \boldsymbol{\pi}_L]$ that is row-wise stacked together to be generated by the mixture TF matrix $\mathcal{D}_{\mathbf{\Psi}}$ at all (t, ω).

The further solution of (24) is the sparse representation of the TF feature vector $\mathcal{D}_{\mathbf{\Pi}}$, that will later construct the estimation of the coefficient matrix in the TF domain.

$$\mathcal{J}(\mathbf{c}_i, \eta) = \frac{1}{2}\|\boldsymbol{\pi}_i - \mathcal{D}_{\mathbf{\Pi}} \mathbf{c}_i\|_2^2 + \eta \|\mathbf{c}_i\|_1, \quad \text{s.t., } \mathbf{c}_{ii} = 0, \tag{24}$$

where $\eta > 0$ is a scalar parameter to balance the trade-off between the sparsity and reconstruction error. $\mathbf{c}_i$ is the corresponding sparse coefficient for $\boldsymbol{\pi}_i$. The maximum value in $\mathbf{c}_i$ indicates the estimated element of $\mathbf{W}$ that is corresponding to $\mathcal{D}_{\mathbf{\Pi}}$. Once a sparse coding problem is built, the solution can be obtained by solving the convex optimization problem.

Here, we use l_1-Homotopy method in [31] to calculate the redundant dictionary $\mathbf{c}_i$ of (24). The procedure obtains a sparse solution with $\mathcal{O}(q^3+L)$ orders, where q is the number of non-zero elements.

D. SOURCE RECOVERY

Since the mixing matrix is not irreversible in the underdetermined BSS [32], the recovered sources also need to be estimated even though the mixing matrix has been known. To obtain a sparse TF representation of the recovered sources, we use the process proposed by [2] with the definition of sub-matrix $\mathcal{W}$ on the following assumption.

Assumption 2: At most $N-1$ sources among M sources are active at each TF point for $M > N$ [8].

Definition 4: Given a matrix $\mathbf{W}$ of size $N \times M$, for any sub-matrices $\mathcal{W}_i$ composed of size $N \times (N-1)$, there are $\binom{M}{N-1}$ possible combinations included in the set $\mathcal{W}$, that is

$$\mathcal{W} = \{\mathcal{W}_i | \mathcal{W}_i = [\boldsymbol{w}_{\lambda_1}, \cdots, \boldsymbol{w}_{\lambda_{N-1}}]\}. \tag{25}$$

Assumption 2 indicates the number of columns of the sub-matrix $\mathcal{W}_i$ to be derived, so that for each TF point (t,ω) we have a corresponding $\mathcal{W}_*$, which satisfies

$$\mathcal{W}_* = \arg \min_{\mathcal{W}_i \in \mathcal{W}} \left\| \mathcal{D}_{\boldsymbol{\Psi}}(t,\omega) - \mathcal{W}_i \mathcal{W}_i^{\dagger} \mathcal{D}_{\boldsymbol{\Psi}}(t,\omega) \right\|_2, \tag{26}$$

where $\mathcal{W}_i^{\dagger}$ is the pseudo-inverse of $\mathcal{W}_i$, which is defined as $\mathcal{W}_i^{\dagger} = (\mathcal{W}_i^{\top}\mathcal{W}_i)^{-1}\mathcal{W}_i^{\top}$.

For a matrix $\mathbf{W}$ of size $N \times M$ $(M > N)$, we want to derive the sub-matrices $\mathcal{W}_i$ of size $N \times M'$, where its columns are excerpted to be independent. Thus, if M' is more than N, the columns of the sub-matrices must be non-independent. There will be exist at least one column vector that can be linearly expressed by other column vectors. Therefore, M' needs to be less than or equal to N. Similar with reference [8], we set the number of columns of sub-matrices as $N-1$, i.e. each $\mathcal{W}_i$ composed of size $N \times (N-1)$, where $M' = N-1$ pick up from total M columns that allow us to compose an optimal sub-matrix $\mathcal{W}_*$ from all possible combinations of the candidate set $\mathcal{W}$, so that (26) is satisfied.

Thus, each source in the TF domain can be estimated by

$$\hat{\mathcal{D}}_{s_j}(t,\omega) = \begin{cases} \mathcal{W}_*^{\dagger}\mathcal{D}_{\boldsymbol{\Psi}}(t,\omega), & \text{if } j = \lambda_i, \\ 0, & \text{otherwise}, \end{cases} \tag{27}$$

where λ_i is the index number of the sub-matrix that implies the non-zero element of $\tilde{\mathcal{D}}_{s_j}$ at each TF point. The source estimator $\tilde{s}_i(t)$ is then obtained by converting $\hat{\mathcal{D}}_{s_i}(t,\omega)$ to the time domain using the inverse STFT.

E. SELECTING FROM THE EXTRACTED COMPONENTS

Due to the multiple subspaces representation, the proposed method forms k extracted components. Therefore, one more thing needs to be considered that is selecting n outputs from k components as the estimator of original sources. We thus use the column-wise singular value decomposition (SVD) to form each column of the original sources $\mathbf{s}$, where the estimator forms all possible k subspaces.

Algorithm 1 Generate Polynomials of Degree 1 by Gram-Schmidt Procedure

Input: N-dimensional observed signals $\mathbf{x}(t) = [x_1(t), \cdots, x_N(t)]^{\top}$.

Output: The recovered signals $\hat{\mathbf{s}}(t) = [s_1(t), \cdots, s_M(t)]^{\top}$ for $t = 1, \cdots, T$.

1: Stage 1:
2: **for** $t = 1 : T$ **do**
3: Mapping the observed signals into multiple spaces $\boldsymbol{\Psi}(\mathbf{x}(t)) = \boldsymbol{\Xi}^{\top}\boldsymbol{\Phi}(\mathbf{x}(t))$.
4: **end for**
5: Stage 2:
6: **for** $i = 1 : k$ **do**
7: Transform $\boldsymbol{\Psi}(t)$ from the time domain into TF domain $\mathcal{D}_{\Psi_i}(\tau,\omega) = \sum_{t=-\infty}^{\infty} \Psi_i(t) h(t-\tau) e^{-j\omega t}$.
8: **end for**
9: To remove the effect of noise, we do $\frac{\|\mathcal{D}_{\boldsymbol{\Psi}}(t_p,\omega_k)\|}{\max_{\omega}\{\|\mathcal{D}_{\boldsymbol{\Psi}}(t_p,\omega)\|\}} > \epsilon$, where $\epsilon = 0.05$ in [8].
10: Minimizing (24) to derive a candidate matrix $\mathbf{W}$ $\mathcal{J}(\mathbf{c}_i,\eta) = \frac{1}{2}\|\boldsymbol{\pi}_i - \mathcal{D}_{\boldsymbol{\Pi}}\mathbf{c}_i\|_2^2 + \eta\|\mathbf{c}_i\|_1$, where $\mathbf{W}$ is formed by the element of $\mathcal{D}_{\boldsymbol{\Pi}}$ that corresponding to the maximum value in $\mathbf{c}_i$.
11: The optimal sub-matrix $\mathcal{W}$ can be derived by (26).
12: Convert the estimated source in the TF domain back to the time domain in (27).
13: Stage 3:
14: **for** $t = 1 : T$ **do**
15: Apply SVD on matrix $\mathbf{F} = [\tilde{s}_1(:,t), \cdots, \tilde{s}_k(:,t)]$.
16: The dominant left singular vector is the estimate of the t-th column of $\hat{s}$, i.e., $\hat{s}(:,t) \leftarrow \mathbf{U}(:,1)$, where $\mathbf{F} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^{\top}$.
17: **end for**

The major steps of the proposed algorithm for multiple subspaces representation are summarized in Algorithm 1. In stage 1: By parameterizing such subspaces, we can map the observed signals in the feature space with the coefficient matrix from the parameter space. In stage 2: We then exploit the linear mixture in the feature space that corresponds to the nonlinear mixture in the input space. Thus, by allowing multiple sources to be presented at any point in the TF domain, we can figure out the target matrix in a sparse mixture of TF vectors. Final stage: Multiple subspaces produce k extracted components $\tilde{\mathbf{s}}$, we need to select n outputs as the estimator of the original sources $\hat{\mathbf{s}}$. Thus, the recovered sources formed from each dominant left singular vector $\mathbf{U}(:,1)$ in the column-wise SVD.

V. EXPERIMENTS AND DISCUSSIONS

To evaluate the proposed algorithm, we performed the simulation on both synthetic data and real audio data over the

underdetermined mixtures. First, using the synthetically generated data, the proposed algorithm is applied to show that the subspace matches the nonlinearity of mixing function in the time domain. Then the nonlinear problem can be separated in the feature space. Next, the recovered sources are tested on two kinds of environment.

A. METHODS AND EVALUATION METRIC

To evaluate the efficiency of the proposed algorithm, we perform a comparison with some developed conventional algorithms, such as the underdetermined BSS (UBSS) method based on the TF non-disjoint assumption [2], the underdetermined convolutive BSS (UCBSS) method[1] based on the subspace representation [33].

The performance of the recovered sources is evaluated by using three kinds of error measure. One is the Pearson correlation coefficient (PCC), which can evaluate the performance for each signal on the definition of

$$\mathrm{PCC}(\mathbf{s}_i, \hat{\mathbf{s}}_i) = \frac{\mathrm{cov}(\mathbf{s}_i, \hat{\mathbf{s}}_i)}{\sigma_{\mathbf{s}_i}\sigma_{\hat{\mathbf{s}}_i}}, \tag{28}$$

where the recovered source and original source are denoted as $\hat{\mathbf{s}}_i$ and $\mathbf{s}_i$, respectively. $\mathrm{cov}(\cdot,\cdot)$ is the covariance between two variables and the standard deviation is denoted as σ.

The normalized mean squared error (NMSE) is another evaluation criterion used to measure the performance on the overall signals, which is defined by

$$\mathrm{NMSE}(\mathbf{s}, \hat{\mathbf{s}}) = 10\log_{10}\left(\frac{1}{M}\sum_{i=1}^{M}\min_{\delta}\frac{\|\mathbf{s}_i - \delta\hat{\mathbf{s}}_i\|_2^2}{\|\mathbf{s}_i\|_2^2}\right). \tag{29}$$

The scalar δ is used for controlling the scalar ambiguity.

During the separation process, the signals may be distorted especially when the sources are overlapped in their TF domain. Hence, it is necessary to measure the distortion and the artifacts introduced by the algorithm to assess the quality of separation. The BSSEVAL toolbox [34] is available online.[2] Then the source-to-distortion ratio (SDR), the source-to-interference ration (SIR), and the source-to-artifacts ratio (SAR) of an estimated source $\hat{s}_{ij}$ as

$$\mathrm{SDR}_j = 10\log_{10}\frac{\sum_{i=1}^{M}\sum_t s_{ij}(t)^2}{\sum_{i=1}^{M}\sum_t [e_{ij}^{\mathrm{spat}}(t) + e_{ij}^{\mathrm{interf}}(t) + e_{ij}^{\mathrm{artif}}(t)]^2},$$

$$\mathrm{SIR}_j = 10\log_{10}\frac{\sum_{i=1}^{M}\sum_t [s_{ij}(t)^2 + e_{ij}^{\mathrm{spat}}(t)]^2}{\sum_{i=1}^{M}\sum_t e_{ij}^{\mathrm{interf}}(t)^2},$$

$$\mathrm{SAR}_j = 10\log_{10}\frac{\sum_{i=1}^{M}\sum_t [s_{ij}(t) + e_{ij}^{\mathrm{spat}}(t) + e_{ij}^{\mathrm{interf}}(t)]^2}{\sum_{i=1}^{M}\sum_t e_{ij}^{\mathrm{artif}}(t)^2},$$

where $\hat{s}_{ij}(t) = s_{ij}(t) + e_{ij}^{\mathrm{spat}}(t) + e_{ij}^{\mathrm{interf}}(t) + e_{ij}^{\mathrm{artif}}(t)$, s_{ij} is the target source with allowed deformation such as filtering or gain, $e_{ij}^{\mathrm{spat}}(t)$ distinct error components representing spatial distortion, $e_{ij}^{\mathrm{interf}}(t)$ accounts for the interference due to unwanted sources, and $e_{ij}^{\mathrm{artif}}(t)$ corresponds to the artifacts introduced by the separation algorithm.

[1] https://slsp.kaist.ac.kr/xe/index.php?mid=software

[2] http://bass-db.gforge.inria.fr/bss_eval

B. THE EFFECT OF MULTI-SUBSPACE REPRESENTATION

To see the effect of multi-subspace representation, we need to show that the subspace is extracted to approximate the varieties of nonlinearity or distortion. First, let us consider the case where the mixture signals $\mathbf{x}$ plotted in Fig. 2(b) are a nonlinear mixture from two sinusoidal signals, which is also used in [19] and [35] with the form of

$$\begin{aligned} x_1(t) &= \exp(s_1(t)) - \exp(s_2(t)), \\ x_2(t) &= \exp(-s_1(t)) + \exp(-s_2(t)), \end{aligned} \tag{30}$$

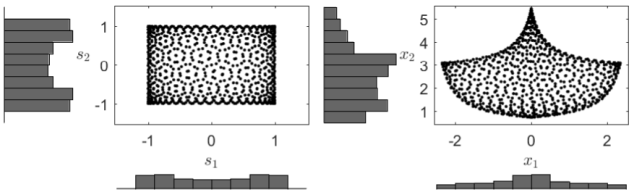

FIGURE 2. An illustration of nonlinear mapping. (a) Original signals generated from two sinusoidal functions. (b) Mixture signals are modeled nonlinearly from (30).

where $s_1(t) = \sin(0.05\pi t)$ and $s_2(t) = \sin(0.021\pi t)$ with the different frequencies. Each source has 1, 000 data points. We indicate the polynomial function of the degree 9 as a kernel function, i.e., $\mathcal{K}(\mathbf{s}_1, \mathbf{s}_2) = (\mathbf{s}_1^\top \mathbf{s}_2 + 1)^9$. Without loss of generality, we further discuss the effect of the different kernel functions. The dimensionality of subspace is set as 20.

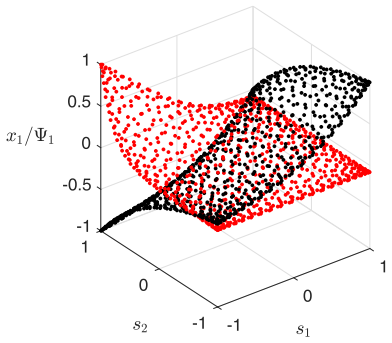

FIGURE 3. The nonlinear mixing $\mathbf{x}_1$ and the subspace constructed by approximation function. The black points illustrate the observed signal $\mathbf{x}_1$ in nonlinear mixing. The red points structure the subspace of best-matching. By using a coefficient matrix, the subspace can be rotated and scaled to match the nonlinear transformation.

As shown in Fig. 3, the nonlinearity of the mixed signals $\mathbf{x}_1$ is comparatively strong that is plotted by black points. The observed data $\mathbf{x}_1$ is first implicitly mapped into feature space, and the effective subspace plotted by red points. Using the coefficient matrix, we can rotate and scale the subspace to match the nonlinear transformation. Relying on this effective subspace, the nonlinear problem can be linearly separable in the feature space, i.e., the original sources can be estimated linearly in the feature space by (20).

One of the keys in the algorithm is to find a set of orthogonal basis to study the parameterized signals in multiple subspaces. Some techniques can help that are roughly analogous in [36]–[38]. To perform the comparison, we employed some classical methods to extract a set of basis in the proposed algorithm, such as kernel principle component analysis (KPCA) [39], k-means [40], and random sampling. To reduce the random effect, 40 times of Monte Carlo simulations are performed.

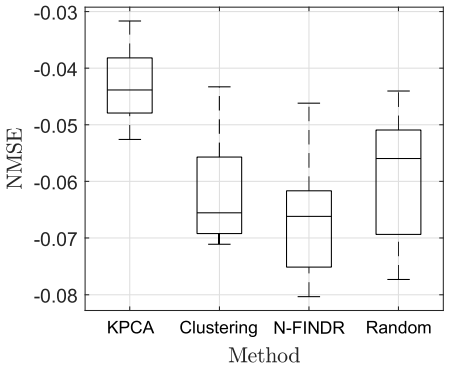

FIGURE 4. The averaged NMSEs of estimators using the different method to form a set of basis.

As we can see in Fig. 4, using N-FINDR to extract a set of basis provides the smaller construction error in the nonlinear approximation. Either KPCA or k-means clustering is considered to obtain some vectors, which is expected to be independent. However, the method may not be appropriate for mixture data. This is due to the independence of the vectors, which can not be guaranteed the mutually orthogonal vectors among the basis. For further work, a set of orthonormal subspaces produced by these k vectors can represent the nonlinearity of mixing functions in the reduced data.

C. SEPARATION OF SPEECH AND AUDIO SIGNALS

To show the separation of speech and audio signals over the undertermined mixtures, the experiments are designed on two kinds of environment. Both cases use the audio data from real-world that are available in the literature [2] and online repositories.[3] The simulation is performed on the following parameter setup, where the proposed method considers the case where some examples of vector dot-product kernel. The dimensionality of subspace is set as 20. The parameter η of scalar regularization is taken as 0.001. Assume that the noise is generated from white and Gaussian with some uncorrelated data points whose variance is usually assumed to be uniform. To reduce the random effect, the simulation is repeated 20 times. The experimental conditions are summarized in TABLE 1.

The first example assumes that the mixture signals are mixed nonlinearly. The mixing functions are employed to transform $m = 4$ independent speech signals for $n = 3$ observations that are available from the literature [2], where

[3]http://bass-db.gforge.inria.fr/BASS-dB/

TABLE 1. The experimental conditions.

Parameters	Values
Sampling rate	8 kHz
Number of sample points	15000 points
Window function	Hanning window
STFT frame size	1024 points (128ms)
Time frame shift	256 points (32ms)

each observation is a linear mixture of nonlinear distorted sources, i.e., $\mathbf{x}(t) = \mathbf{A}\exp(\mathbf{s}(t))$. Here, the exponential transformation provides a nonlinear distortion and the matrix $\mathbf{A}$ randomly generated from a uniform distribution $U[-1, 1]$. Since there is no good path to choose a kernel function, unless we have some prior information about the data that might be helpful to determine a proper kernel function [41]. Here, we only consider the kernel function with 3 classical types, where polynomial kernel of degree 9 is given by $\mathcal{K}(\mathbf{x}, \mathbf{y}) = (\mathbf{x}^\top\mathbf{y} + 1)^9$, Radial-basis function (RBF) of uniform variance has the definition of $\mathcal{K}(\mathbf{x}, \mathbf{y}) = \exp(-\frac{\|\mathbf{x}-\mathbf{y}\|^2}{2})$, and sigmoid function is formed as $\mathcal{K}(\mathbf{x}, \mathbf{y}) = \tanh(\mathbf{x}^\top\mathbf{y})$, respectively. The results are given under the signal-to-noise power ratio (SNR) in the range of 5 dB to 45 dB. The experiments are repeated 20 times.

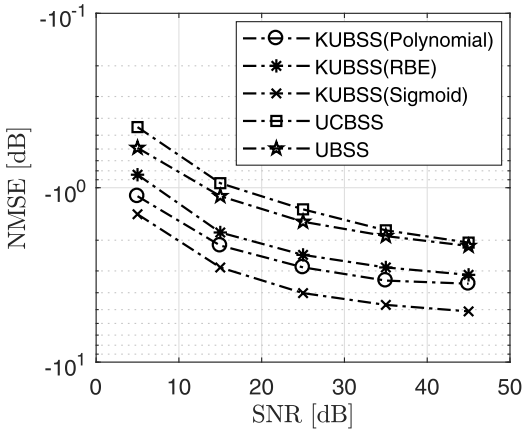

FIGURE 5. The averaged NMSEs on the different SNR levels. Here the number of sources $M = 4$ and that of observations $N = 3$.

In Fig. 5, the separation accuracy is compared with some conventional algorithms on the different SNR levels. We can see that the proposed kernel-based underdetermined blind source separation (KUBSS) algorithm consistently provides a higher accuracy over the whole SNR range. When the SNR reaches 25 dB, NMSEs decrease linearly with further increasing of SNR. Benefiting from a multi-subspace representation, the effective subspace can extract the nonlinearity or distortion caused by nonlinear mixing in kernel feature space. Moreover, this is because both UBSS and UCBSS methods are based on single source detection, which is built on the assumption that there exists only a single source or

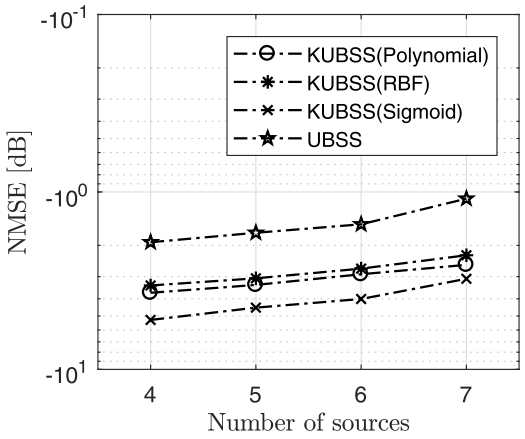

FIGURE 6. The averaged NMSEs on the number of sources increases from $M = 4$ to 7.

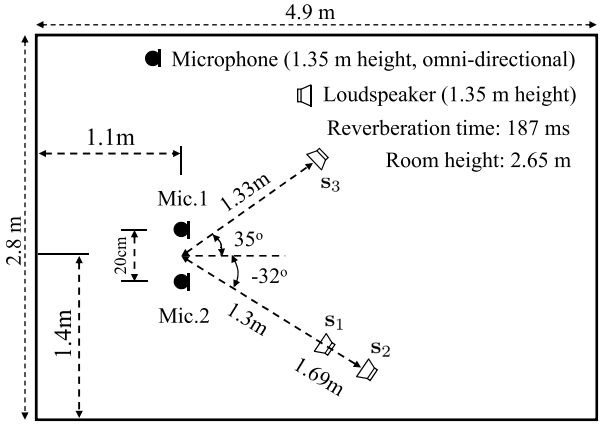

FIGURE 7. The virtual room environment for synthetic mixtures.

dominant energy of its corresponding single source at the TF points.

Experiment 2 shows NMSEs of the proposed algorithm where the observations are generated from the enhancement of the undetermined level, i.e., the number of sources is increased from 4 to 7 while that of observations is kept as 3. In general, a larger number of observations leads to better separation accuracy. The NMSE improvements for different combinations of sources and observations are shown in Fig. 6, where a set of basis is extracted using the N-FINDR approach. The kernel function also works on 3 types and 20 experiments are repeated.

Fig. 6 illustrates the averaged NMSEs when the number of sources increases from $M = 4$ to 7. The proposed algorithm with the "RBF" function achieved about 1.5 dB higher NMSEs against other algorithms over the whole range. In addition, 3.2 dB higher NMSEs are shown than the other algorithms when we use "Sigmoid" function. However, the performance degraded as the number of the underlying sources increased. In practice, this is due to the fact that the sources are not perfectly disjoint in the TF domain [42], which leads to the estimation error of

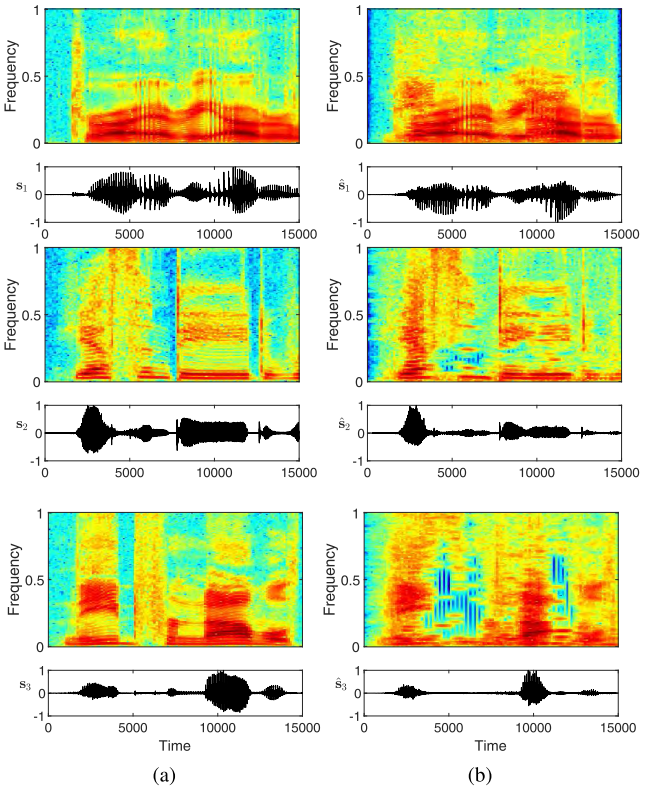

FIGURE 8. The spectrogram of signals with three channels. (a) The three subfigures represent the original sources of s_1, s_2, and $s_3$iijN respectively. (b) The three subfigures correspond to the recovered sources of $\hat{s}_1$, $\hat{s}_2$, and $\hat{s}_3$, respectively.

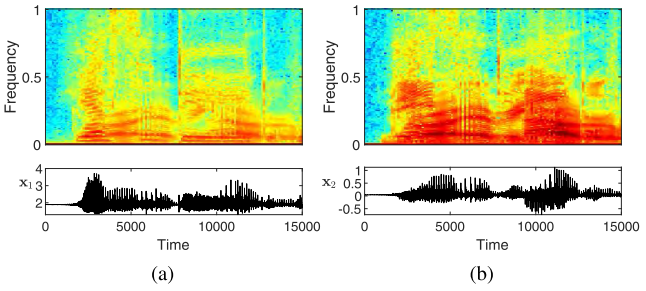

FIGURE 9. The mixture is achieved by transforming 3 original sources to 2 observations. The mixed signals x_1 and x_2 are shown in the (a) and (b), respectively.

recovered signals. As the number of sources increases, the overlap will occur in the spectra as well as the estimation error also increase.

D. EXPERIMENTS USING REAL ROOM IMPULSE RESPONSES

The experiments were designed on speech data with impulse responses in an office room. The observations are collected from this room with 187 ms reverberation time. The effect of the impulse response is measured in the face of using "Sample Champion" software that is available online.[4]

[4]http://www.purebits.com

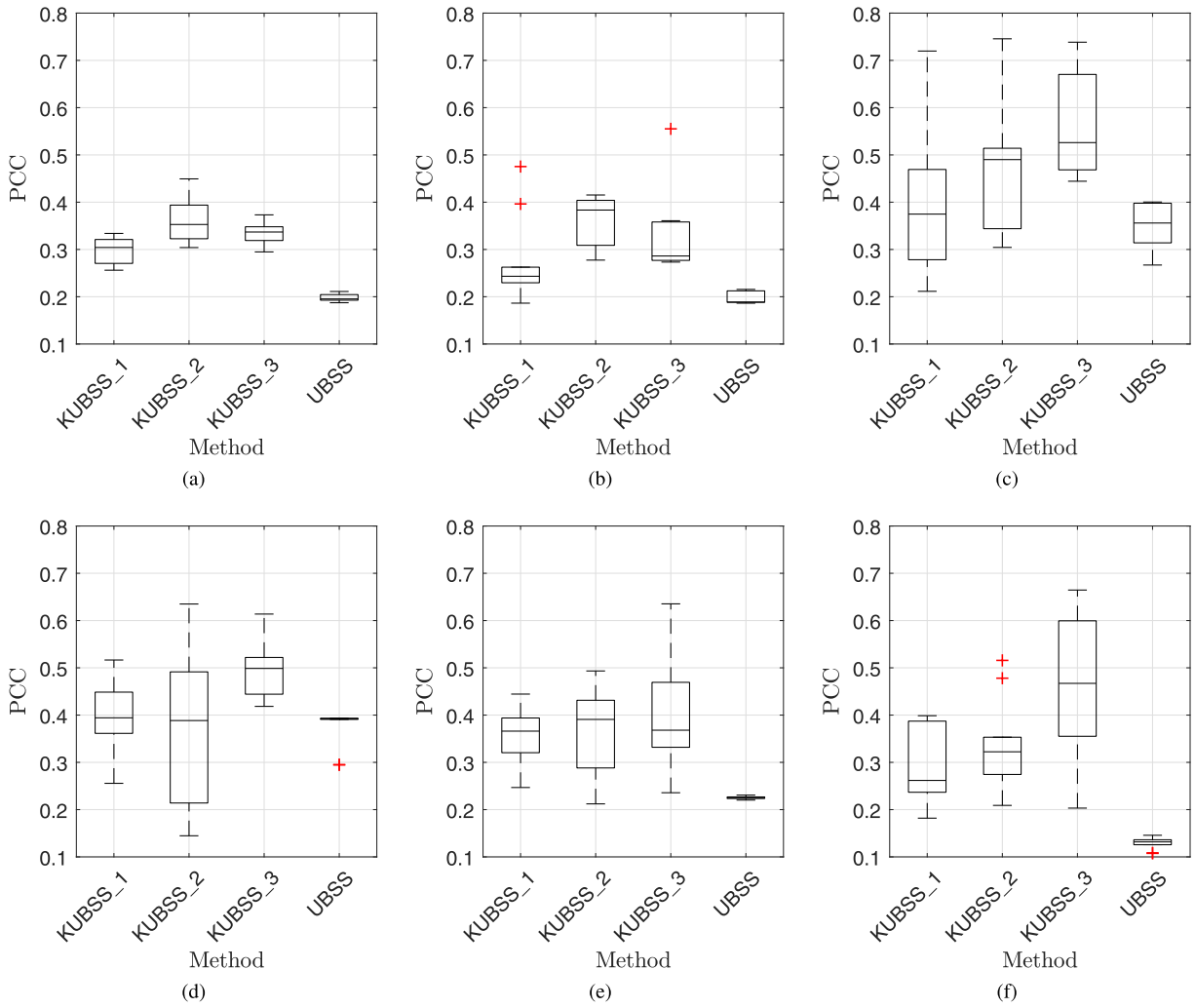

FIGURE 10. Separation of the speech data with impulse responses. The first column (a)(d) are the results from the collinear mixture of $\mathbf{s}_1$ and $\mathbf{s}_2$. The results of the non-collinear mixture are shown, respectively, in the middle column (b)(e) of $\mathbf{s}_1$ and $\mathbf{s}_3$ mixture, and the third column (c)(f) of $\mathbf{s}_2$ and $\mathbf{s}_3$ mixture. The first row (a)-(c) are PCCs of the estimated signal $\hat{\mathbf{s}}_1$. The second row (d)-(f) are PCCs of the estimated signal $\hat{\mathbf{s}}_2$.

Fig. 8 shows the original sources $\mathbf{s}(t)$ of 3 channels. Without loss of generality, the microphone, and loud speaker transfer function is neglected in the measurements [42]. The virtual room environment is illustrated in Fig. 7. A two-element microphone array was used for recording speech signals, which arrived in two different directions, such as 35° and −32°. It is worth noting that the source $\mathbf{s}_1$ and $\mathbf{s}_2$ are collinear that provides a challenging task using independent component analysis. The underdetermined mixture is achieved by transforming 3 original sources $\mathbf{x}(t)$ to 2 observations that are given in Fig. 9.

The experiments involve three scenarios, where the first case is a collinear mixture, i.e., mixed signals generated from sources $\mathbf{s}_1$ and $\mathbf{s}_2$. The second case is considered by a non-collinear mixture from $\mathbf{s}_1$ and $\mathbf{s}_3$, or $\mathbf{s}_2$ and $\mathbf{s}_3$. The third case is underdetermined mixture using all the three sources, i.e., $\mathbf{s}_1$, $\mathbf{s}_2$, and $\mathbf{s}_3$ in Fig. 7. Also, 3 classical kernel functions are used for comparison, such as "polynomial kernel", "RBF kernel", and "Sigmoid kernel". In the legend of the figure, they are denoted as "KUBSS_1", "KUBSS_2", and "KUBSS_3", respectively, for convenience. The Pearson correlation coefficient (PCC) is used to evaluate the performance of each signal.

From Fig. 10, it can be seen that the algorithms can recover the original sources in all the 3 cases. We further show PCC of each channel between the original source and the recovered source using the PCC (28) measure. As shown in the figures, the proposed approach exhibits the promising results. This is due to the fact that the UBSS algorithm is lack of analysis of nonlinearity. In addition, the average SDR, SIR, and SAR are adopted as a performance measure of the source recovery. The performance shown in TABLE 2 are mean performances of 20 experiments. As we can see, the proposed algorithm performed better in terms of average SDR, SIR, and SAR compared with that of the UBSS methods. One can notice that the collinear mixture provides a lower accuracy than

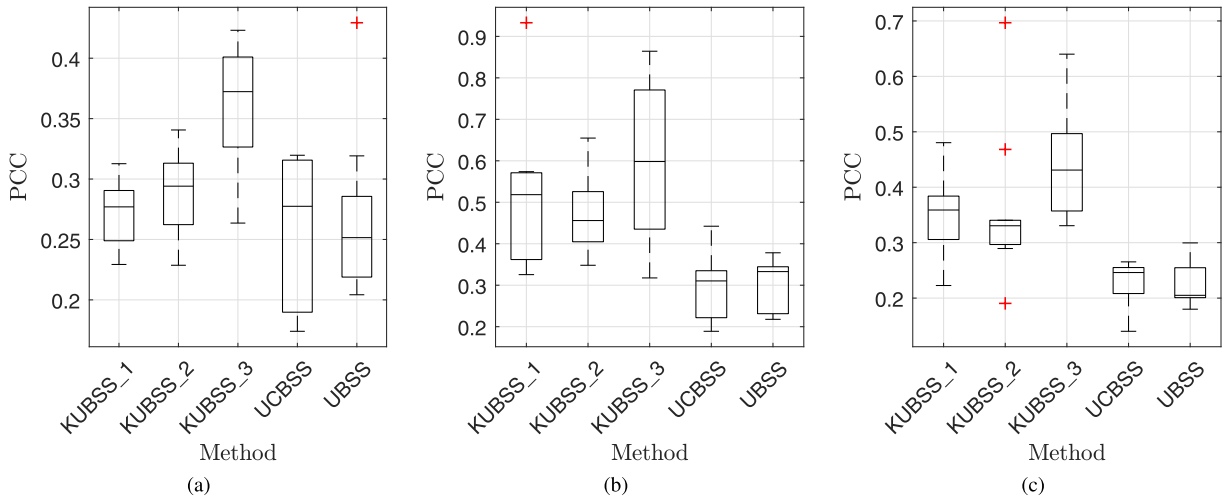

FIGURE 11. Separation of the speech data on the underdetermined mixture with the impulse response. (a) is the performance of estimated signal $\hat{s}_1$, (b) is the performance of estimated signal $\hat{s}_2$, and (c) corresponds to the estimated signal $\hat{s}_3$.

TABLE 2. Performance comparison of the proposed algorithm, where the algorithm UCBSS [33] only works on the underdetermined mixture.

Active sources	Performance measure	Methods				
		KUBSS_1	KUBSS_2	KUBSS_3	UCBSS	UBSS
$\mathbf{s}_1$ and $\mathbf{s}_2$ (Collinear)	SDR	1.86	2.11	2.37	–	1.13
	SIR	2.24	2.49	4.69	–	3.75
	SAR	6.97	6.56	6.90	–	7.48
$\mathbf{s}_1$ and $\mathbf{s}_3$ (Non-collinear)	SDR	4.26	4.31	2.74	–	2.61
	SIR	4.01	6.83	2.72	–	2.73
	SAR	6.68	8.59	5.92	–	6.10
$\mathbf{s}_2$ and $\mathbf{s}_3$ (Non-collinear)	SDR	2.25	2.44	2.77	–	2.59
	SIR	4.09	3.87	5.59	–	4.02
	SAR	5.44	6.62	7.17	–	6.63
$\mathbf{s}_1$, $\mathbf{s}_2$, and $\mathbf{s}_3$ (Underdetermined)	SDR	6.97	6.64	6.64	4.61	3.60
	SIR	4.93	6.22	6.21	4.49	2.47
	SAR	4.58	6.72	6.73	6.72	5.57

non-collinear mixture on speech sources. Therefore, a large enough angle between two sources is a crucial condition to obtain good separation performance. Some discussions have been studied in [43]. The limitation is not only for our study, but also the limitation of the separation filter obtained by ICA that forms spatial directivity [44].

Furthermore, Fig. 11 shows the averaged PCC on the underdetermined mixture with the impulse response. As we can see, despite adopting a similar assumption to extract sources, the proposed method exhibits a high separation accuracy compared with that of the UCBSS and UBSS methods. The main reason is that subspaces can extract the nonlinearity caused by the mixing function. As shown in TABLE 2, the proposed algorithm performs better in terms of average SDR, SIR, and SAR for situations tested. The coefficient matrix is estimated by minimizing the cost function, which is directly related to the evaluation criterion. In addition, the compared methods always require the sparsity of the sources to some extent, while the assumption may not be satisfied in reality.

VI. CONCLUSIONS

In this paper, we propose a multi-subspace representation based separation approach that tackles the scenario of the nonlinear and underdetermined mixture. The separation system is constructed using the kernel methods with a multi-subspace structure. One of the keys in that algorithm is to find a set of orthogonal basis to study the parameterized signals in multiple feature spaces. We attempt to use the geometric vertices of the convex hull as the basis, which parameterizes the multi-subspace that contains the reduced vectors in the feature space. Relying on a set of an orthonormal basis, the spanned subspaces can represent the nonlinearity of mixing function in the minimum number.

Another contribution is to derive the coefficient matrix by solving an optimization problem on the coding

coefficient vector. Once such subspaces are built, by allowing multiple sources to be presented at any point in the TF domain, we can figure out the target matrix in sparse mixture TF vectors with less computational cost. Finally, using this coefficient matrix, the original sources in underdetermined scenarios can be estimated. The experiments are designed on two kinds of environment, such as the signals perform nonlinear mixing, or mixing with some direction angles in a virtual room environment. The proposed approach exhibits a higher separation accuracy than that of the conventional algorithms.

REFERENCES

[1] 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.

[2] L. Zhen, D. Peng, Z. Yi, Y. Xiang, and P. Chen, "Underdetermined blind source separation using sparse coding," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 12, pp. 3102–3108, Dec. 2017.

[3] Z. Liang, C. Xun, X. Ji, and Z. J. Wang, "Underdetermined joint blind source separation of multiple datasets," *IEEE Access*, vol. 5, pp. 7474–7487, 2017.

[4] 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.

[5] A. Belouchrani, M. G. Amin, N. Thirion-Moreau, and Y. D. Zhang, "Source separation and localization using time-frequency distributions: An overview," *IEEE Signal Process. Mag.*, vol. 30, no. 6, pp. 97–107, Nov. 2013.

[6] 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.*, Jun. 2000, pp. 2985–2988.

[7] Ö. 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.

[8] A. Aïssa-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.

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

[10] C. Jutten and J. Karhunen, "Advances in nonlinear blind source separation," in *Proc. 4th Int. Symp. Independ. Compon. Anal. Blind Signal Separat.*, Nara, Japan, Apr. 2003, pp. 245–256.

[11] L. Yang, Y. Xiang, and D. Peng, "Precoding-based blind separation of MIMO FIR mixtures," *IEEE Access*, vol. 5, pp. 12417–12427, 2017.

[12] R. Liu, X. Zhu, Y. Jiang, X. Dong, and F. Zheng, "Blind PAPR reduction and ICA based equalization for mmWave FBMC-OQAM systems," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2018, pp. 1–6.

[13] L. Wang and T. Ohtsuki, "Polynomial networks representation of nonlinear mixtures with application in underdetermined blind source separation," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2019, pp. 3687–3691.

[14] M. Banuelos, S. Sindi, and R. F. Marcia, "Negative binomial optimization for biomedical structural variant signal reconstruction," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Apr. 2018, pp. 906–910.

[15] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*. New York, NY, USA: McGraw-Hill, 1991. [Online]. Available: https://dcetit.files.wordpress.com/2013/03/papoulis-probability-random-variables-and-stochastic-processes.pdf

[16] A. Hyvärinen and P. Pajunen, "Nonlinear independent component analysis: Existence and uniqueness results," *Neural Netw.*, vol. 12, no. 3, pp. 429–439, Apr. 1999.

[17] G. Burel, "Blind separation of sources: A nonlinear neural algorithm," *Neural Netw.*, vol. 5, no. 6, pp. 937–947, Nov./Dec. 1992.

[18] Y. Tan, J. Wang, and J. M. Zurada, "Nonlinear blind source separation using a radial basis function network," *IEEE Trans. Neural Netw.*, vol. 12, no. 1, pp. 124–134, Jan. 2001.

[19] S. Harmeling, A. Ziehe, M. Kawanabe, and K.-R. Müller, "Kernel-based nonlinear blind source separation," *Neural Comput.*, vol. 15, no. 5, pp. 1089–1124, May 2003.

[20] D. Martinez and A. Bray, "Nonlinear blind source separation using kernels," *IEEE Trans. Neural Netw.*, vol. 14, no. 1, pp. 228–235, Jan. 2003.

[21] Z. Wang, K. Crammer, and S. Vucetic, "Multi-class pegasos on a budget," in *Proc. 27th Int. Conf. Mach. Learn. (ICML)*, Jun. 2010, pp. 1143–1150.

[22] L. Wang and T. Ohtsuki, "Nonlinear blind source separation unifying vanishing component analysis and temporal structure," *IEEE Access*, vol. 6, pp. 42837–42850, 2018.

[23] L. Wang and T. Ohtsuki, "Signal restoration based on temporal structure and multi-layer architecture," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2018, pp. 1–7.

[24] S. Harmeling, A. Ziehe, M. Kawanabe, B. Blankertz, and K.-R. Müller, "Nonlinear blind source separation using kernel feature spaces," in *Proc. Int. Workshop Independ. Compon. Anal. Blind Signal Separat.*, Dec. 2001, pp. 102–107.

[25] S. Harmeling, A. Ziehe, M. Kawanabe, and K.-R. Müller, "Kernel feature spaces and nonlinear blind souce separation," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, vol. 14, Dec. 2002, pp. 761–768.

[26] S. Boyd and L. Vandenberghe, *Convex optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2009.

[27] A. Ambikapathi, T.-H. Chan, W.-K. Ma, and C.-Y. Chi, "Chance-constrained robust minimum-volume enclosing simplex algorithm for hyperspectral unmixing," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 11, pp. 4194–4209, Nov. 2011.

[28] M. E. Winter, "N-FINDR: An algorithm for fast autonomous spectral end-member determination in hyperspectral data," *Proc. SPIE*, Oct. 1999, pp. 266–277.

[29] C.-H. Lin, C.-Y. Chi, Y.-H. Wang, and T.-H. Chan, "A fast hyperplane-based minimum-volume enclosing simplex algorithm for blind hyperspectral unmixing," *IEEE Trans. Signal Process.*, vol. SP-64, no. 8, pp. 1946–1961, Apr. 2015.

[30] A. Taleb and C. Jutten, "Source separation in post-nonlinear mixtures," *IEEE Trans. Signal Process.*, vol. 47, no. 10, pp. 2807–2820, Oct. 1999.

[31] M. S. Asif and J. Romberg, "Sparse recovery of streaming signals using ℓ_1-Homotopy," *IEEE Trans. Signal Process.*, vol. 62, no. 16, pp. 4209–4223, Aug. 2014.

[32] 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.

[33] J. Cho and C. D. Yoo, "Underdetermined convolutive BSS: Bayes risk minimization based on a mixture of super-Gaussian posterior approximation," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 23, no. 5, pp. 828–839, May 2015.

[34] E. Vincent, S. Araki, F. Theis, G. Nolte, P. Bofill, H. Sawada, A. Ozerov, V. Gowreesunker, D. Lutter, N. Q. K. Duong, "The signal separation evaluation campaign (2007–2010): Achievements and remaining challenges," *Signal Process.*, no. 92, pp. 1928–1936, Aug. 2012.

[35] L. Wang and T. Ohtsuki, "Underdetermined blind separation using multi-subspace representation in time-frequency domain," in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2019.

[36] J. B. Tenenbaum, V. De Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, no. 5500, pp. 2319–2322, Dec. 2000.

[37] Z. Zhang and H. Zha, "Principal manifolds and nonlinear dimensionality reduction via tangent space alignment," *SIAM J. Sci. Comput.*, vol. 26, no. 1, pp. 313–338, 2004.

[38] S. Yan, D. Xu, B. Zhang, H.-J. Zhang, Q. Yang, and S. Lin, "Graph embedding and extensions: A general framework for dimensionality reduction," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 1, pp. 40–51, Jan. 2007.

[39] B. Schölkopf, A. Smola, and K.-R. Müller, "Kernel principal component analysis," in *Proc. Int. Conf. Artif. Neural Netw.*, Jun. 1997, pp. 583–588.

[40] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York, NY, USA: Springer, 2006.

[41] B. W. Silverman, *Density Estimation for Statistics and Data Analysis*. London, U.K.: Chapman-Hall, 1986.

[42] V. G. Reju, S. N. Koh, and Y. Soon, "Underdetermined convolutive blind source separation via time–frequency masking," *IEEE Trans. Audio, Speech, Language Process.*, vol. 18, no. 1, pp. 101–116, Jan. 2010.

[43] V. G. Reju, S. N. Koh, and Y. Soon, "Partial separation method for solving permutation problem in frequency domain blind source separation of speech signals," *Neurocomputing*, vol. 71, nos. 10–12, pp. 2098–2112, Jun. 2008.
[44] R. Mukai, H. Sawada, S. Araki, and S. Makino, "Frequency-domain blind source separation of many speech signals using near-field and far-field models," *EURASIP J. Adv. Signal Process.*, Dec. 2006, Art. no. 083683.

LU WANG received the B.E. and M.E. degrees from the Faculty of Electrical Engineering, Heilongjiang University, Harbin, China, in 2008 and 2011, respectively, and the Ph.D. degree from the Institute of Scientific and Industrial Research, Osaka University, Japan, in 2016. She is currently pursuing the Ph.D. degree with the Graduate School of Science and Technology, Keio University, Yokohama, Japan. From 2016 to 2017, she was a Research Associate with the Department of Reasoning for Intelligence, Osaka University, Japan. Her research interests include blind source separation, model mining, and knowledge discovery from massive data with a recent emphasis on improvement of noisy speech and biomedical engineering applications.

TOMOAKI OHTSUKI (SM'01) received the B.E., M.E., and Ph.D. degrees in electrical engineering from Keio University, Yokohama, Japan, in 1990, 1992, and 1994, respectively, where he was a Postdoctoral Fellow and a Visiting Researcher in electrical engineering, from 1994 to 1995. From 1993 to 1995, he was a Special Researcher of Fellowships of the Japan Society for the Promotion of Science for Japanese Junior Scientists. From 1995 to 2005, he was with the Science University of Tokyo. In 2005, he joined Keio University, where he is currently a Professor. From 1998 to 1999, he was with the Department of Electrical Engineering and Computer Sciences, University of California at Berkeley, Berkeley. He has published more than 160 journal papers and 370 international conference papers. His research interests include wireless communications, optical communications, signal processing, and information theory.

He is a Fellow of the IEICE. He was a recipient of the 1997 Inoue Research Award for Young Scientist, the 1997 Hiroshi Ando Memorial Young Engineering Award, the Ericsson Young Scientist Award 2000, the IEEE the 1st Asia-Pacific Young Researcher Award 2001, the 2002 Funai Information and Science Award for Young Scientist, the 5th International Communication Foundation (ICF) Research Award, the 2011 IEEE SPCE Outstanding Service Award, the 27th TELECOM System Technology Award, ETRI Journal's 2012 Best Reviewer Award, and the 9th International Conference on Communications and Networking in China 2014 (CHINACOM) Best Paper Award. He served a Chair of the IEEE Communications Society, Signal Processing for Communications and Electronics Technical Committee. He has served as a General-Co Chair and a Symposium Co-Chair for many conferences, including IEEE GLOBECOM 2008, SPC, IEEE ICC2011, CTS, IEEE GCOM2012, SPC, and IEEE SPAWC. He served as a Technical Editor for the *IEEE Wireless Communications Magazine* and an Editor for Elsevier *Physical Communications*. He is currently serving as an Area Editor for the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY and an Editor for the IEEE COMMUNICATIONS SURVEYS AND TUTORIALS. He gave tutorials and keynote speech at many international conferences, including IEEE VTC, IEEE PIMRC, and so on. He was the Vice President of the Communications Society of the IEICE.