Dependent Component Analysis: Concepts and Main Algorithms

Rui Li
School of Sciences, Henan University of Technology, Zhengzhou 450052, P.R.China
Email: liruilyric@126.com

Hongwei Li♣ and Fasong Wang
School of Mathematics and Physics, China University of Geosciences, Wuhan 430074, P.R.China
Email: hwli@cug.edu.cn, fasongwang@126.com

***Abstract*—Dependent Component Analysis(DCA) as an extension of Independent Component Analysis(ICA) for Blind Source Separation(BSS) has more applications than ICA and received more and more attentions during the last several years in the study of signal processing and neural networks. After a general and detailed definition of the DCA model is given, the separateness and uniqueness of the DCA model have been discussed in theory. Then, the state-of-art DCA algorithms are overviewed, these methods include multidimensional ICA, variance dependent BSS, subband decomposition ICA, maximum non-Gaussianity method, Wold decomposition method and time-frequency method are constructed for the BSS problem in theories and some simulations of these algorithms are also exhibited for different applications.**

***Index Terms*—Dependent Component Analysis(DCA), Blind Source Separation(BSS), Independent Component Analysis(ICA); Neural Network; Sparse Component Analysis(SCA)**

I. INTRODUCTION

Blind source separation(BSS) methods have been successfully applied to many areas of science[1,2]. The basic model assumes that the observed signals are linear super-positions of underlying hidden source signals. Let us denote the n source signals by the vector $\mathbf{s}(t)=(s_1(t),\cdots,s_n(t))^T$, and the observed signals by $\mathbf{x}(t)=(x_1(t),\cdots,x_m(t))^T$. Now the mixing can be expressed as

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

where the matrix $\mathbf{A}=[a_{ij}]\in R^{m\times n}$ collects the mixing coefficients. No particular assumptions on the mixing coefficients are made. However, some weak structural assumptions are often made: for example, it is typically assumed that the mixing matrix is square, that is, the number of source signals equals the number of observed signals ($m=n$), the mixing process $\mathbf{A}$ is defined by an even-determined (i.e. square) matrix and, provided that it is non-singular, the underlying sources can be estimated by a linear transformation, which we will assume here as well. If $m>n$, the mixing process $\mathbf{A}$ is defined by an over-determined matrix and, provided that it is full rank, the underlying sources can be estimated by least-squares optimization or linear transformation involving matrix pseudo-inversion. If $m<n$, then the mixing process is defined by an under-determined matrix and consequently source estimation becomes more involved and is usually achieved by some non-linear technique. For technical simplicity, we shall also assume that all the signals have zero mean, but this is no restriction since it simply means that the signals have been centered. $\mathbf{n}(t)=(n_1(t),\cdots,n_m(t))^T$ is a vector of additive noise that is assumed to be zero in this paper. The problem of BSS is now to estimate both the source signals $s_j(t)$ and the mixing matrix $\mathbf{A}$ based on observations of the $x_i(t)$ alone [1-2].

In most BSS methods, the source signals are assumed to be statistically independent. BSS based on such a model is called independent component analysis (ICA). By using non-Gaussianity of the sources, the mixing matrix can be estimated and the source signals can be extracted under appropriate conditions. There are also further approaches of BSS, for example, based on the time structure of the ICs that allow for the estimation of the model. These assumptions are alternatives to the assumption of nongaussianity made in many works. First, it assumes that the ICs have different autocovariances (in particular, they are all different from zero). Second, it considers the case where the variances of the ICs are nonstationary. Finally, it discusses the case that the mixing matrix changes in time. Another increasingly popular and powerful assumption is that the sources have a parsimonious representation in a given basis, these methods have come to be known as sparse methods, they can be called as sparse component analysis.

However, the independence property of sources may not hold in some real-world situations, especially in biomedical signal processing and image processing, and

♣ Corresponding author, hwli@cug.edu.cn, fasongwang@126.com

doi:10.4304/jcp.5.4.589-597

therefore the standard ICA cannot give the expected results. Among many extensions of the basic ICA models, several researchers have studied the case where the source signals are not statistical independent, we call these models dependent component analysis(DCA) model as a whole. The first extended ICA model is the Multidimensional Independent Component Analysis (MICA) model[3], which is a linear generative model as in equation (1). In contrast to ordinary ICA, however, the components (responses) are not assumed to be all mutually independent. Instead, it is assumed that the source signals can be divided into couples, triplets, or in general *i*-tuples, such that the source signals inside a given *i*-tuple may be dependent on each other, but dependencies among different *i*-tuples are not allowed. Based on this basic extension of the ICA model, there have emerged lots of DCA models and corresponding algorithms, such as independent subspace analysis[4], variance dependent BSS[5-7], topographic ICA[8], and tree-dependent component analysis[9], subband decomposition ICA(SDICA)[10-13], maximum non-Gaussianity method [14-16], spectral decomposition method[17], time-frequency method[18-19].

This paper is organized as follows: Section 2 introduces the basic ICA and DCA BSS models and the corresponding uniqueness and indeterminacies in detail; Then in section 3, we describe the state-of-art DCA models and methods for dependent sources in detail; Finally, section 4 concludes the paper.

II. Basic DCA Model and Relationship with ICA

In this section, we will first review the basic ICA model and its indeterminacy, then based on this primary model, the basic DCA model is discussed in detail and the relationship between them is given too.

A. ICA Model

The noiseless instantaneous ICA model [1-2] can be described as

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) \tag{2}$$

where $\mathbf{s}(t)$, $\mathbf{A}$ and $\mathbf{x}(t)$ are defined as in part I. The following assumptions for the model to be identified are needed:

1) The sources are statistically mutually independent;

2) At most one of the sources has Gaussian distribution (If the source signals are random variables);

3) Mixing matrix $\mathbf{A}$ is column full-rank.

The task of ICA is to recover the original signals from the observations $\mathbf{x}(t)$ without the knowledge of $\mathbf{A}$ nor $\mathbf{s}(t)$. Let us consider a linear feed forward memoryless neural network which maps the observation $\mathbf{x}(t)$ to $\mathbf{y}(t)$ by the following linear transform

$$\mathbf{y}(t) = \mathbf{W}\mathbf{x}(t) = \mathbf{W}\mathbf{A}\mathbf{s}(t),$$

where $\mathbf{W} = [w_{ij}] \in R^{n\times m}$ is a separating matrix, $\mathbf{y}(t) = (y_1(t), \cdots, y_n(t))^T$ is an estimate of the possibly scaled and permutated vector of $\mathbf{s}(t)$ and also the network output signals whose elements are statistically mutually independent, so that the output signals $\mathbf{y}(t)$ are possibly scaled estimation of source signals $\mathbf{s}(t)$.

For the indeterminacy of the model (1) itself, we can only get the separation matrix $\mathbf{W}$ satisfied $\mathbf{W}\mathbf{A} = \mathbf{Q}\mathbf{\Lambda}$, where $\mathbf{\Lambda} = diag\{\lambda_1, \lambda_2, \cdots, \lambda_n\}$ is a nonsingular diagonal matrix and $\mathbf{Q}$ is a permutation matrix. Then

$$\mathbf{y}(t) = \mathbf{W}\mathbf{x}(t) = \mathbf{Q}\mathbf{\Lambda}\mathbf{s}(t) = (\lambda_1 s_{k_1}, \cdots, \lambda_n s_{k_n}).$$

That is $y_i(t) = \lambda_i s_{k_i}(t)$, $1 \le i \le n$ and $(k_1, k_2, \cdots, k_n)$ is a permutation of $(1, 2, \cdots, n)$. So there are two indeterminacies in ICA: (1) scaling ambiguity; (2) permutation ambiguity. But this does not affect the application of ICA, because the main information of the signals is included in the waveform of them. Under this condition, Comon proved that the separated signals $y_i(t)$, $i = 1, 2, \cdots, n$ are mutually independent too [8].

Theorem 1. Let $\mathbf{s}$ be a random vector with independent components, of which at most one is Gaussian. Let $\mathbf{C}$ is an invertible matrix and $\mathbf{y} = \mathbf{C}\mathbf{s}$, then the following two properties are equivalent:

(1) The components y_i are mutually independent;

(2) $\mathbf{C} = \mathbf{Q}\mathbf{\Lambda}$,where $\mathbf{Q}$ is a permutation matrix and $\mathbf{\Lambda}$ is a nonsingular diagonal matrix.

B. DCA Model

Multidimensional BSS is the recovery of underlying sources $\mathbf{s}$ from an observed mixtures $\mathbf{x}$. As usual, $\mathbf{s}$ has to fulfill additional properties such as independence or diagonality of the autocovariances (if $\mathbf{s}$ possesses time structure). However in contrast to ordinary BSS, MBSS is more general as some source signals are allowed to contain statistics except independence. One possible solution for MBSS is MICA-an extended ICA model [3].

We have two approaches to relax ICA model, one is add structure to the sources which contain both temporal and spatial indices; another is relax some assumptions in the model (1). Instead of requiring the source independence conditions to hold among all sources, we only assume that they are valid between groups of them, but we do not assume any assumptions within the groups. This relaxed model is denoted by MICA or group ICA [3].

A random vector $\mathbf{y}$ is called an independent component of the random vector $\mathbf{x}$, if there exists an invertible matrix $\mathbf{A}$ and a decomposition $\mathbf{x} = \mathbf{A}(y_1, \cdots, y_i, \cdots, y_j, \cdots)$ such that y_i and y_j are stochastically independent.

For simplicity, it is often assumed that all groups are of the same size k, and that Nk sources are to be extracted from equally many mixtures. In the case of MICA, MBSS implies that the source random vector $\mathbf{s}$ is now only *k*-independent i.e. that $(s_1, \cdots, s_k)^T$, $\cdots$, $(s_{Nk-k+1}, \cdots, s_{Nk})^T$ is mutually independent.

The idea of MICA is that we do not require full independence of transform $\mathbf{y}(t) = \mathbf{W}\mathbf{x}(t)$ but only

mutual independence of certain tuples $y_{i_1},\cdots,y_{i_k}$. If the size of all tuples is restricted to one, this reduces to general ICA problem. In general, of cause the tuples could have different sizes, but for the sake of simplicity, we assume that all tuples have the same size k (we call it regular MICA). If the model has one tuple only, and the components in the tuple are statistical dependent, the model is the most general DCA model. Next, for simplicity, we give the MICA model definition and its indeterminacy.

Cardoso [3] generalized ICA into MICA from the geometrical view. First give a definition.

Definition 1. Let $E_1,\cdots,E_c$ be c linear subspaces of R^n. They are said to be linearly independent if any vector $\mathbf{x}$ of $E_1\oplus\cdots\oplus E_c$ admits of a unique decomposition as $\mathbf{x}=\sum_{p=1}^{c}\mathbf{x}_p$, with $\mathbf{x}_p\in E_p$ for $1\le p\le c$. In such a case, the vectors $\mathbf{x}_1,\cdots,\mathbf{x}_c$ are called the linear components of $\mathbf{x}$ on the set $E_1,\cdots,E_c$.

Definition 2. A random n-dimensional vector $\mathbf{x}$ admits of a MICA decomposition $\{\mathbf{x}_1,\cdots,\mathbf{x}_c\}$ in c components if it exists c linearly independent 'component subspaces' $E_1,\cdots,E_c$ of R^n on which the linear components of $\mathbf{x}$ are statistically independent.

From definition 2, there may be some ambiguities such as 1) maximality of the decomposition; 2) The Gaussian component, so [3] give a canonical definition of MICA decomposition.

Definition 3. The canonical MICA decomposition(if it exist) of a vector $\mathbf{x}$ is the unique MICA decomposition of $\mathbf{x}$ into $\mathbf{x}=\sum_{p=1}^{c}\mathbf{x}_p$ such that: 1) there is at most one Gaussian component; 2) no non-Gaussian component can be further decomposed into independent components.

III. Extended DCA Models and Methods

Based on the basic DCA model, some extended DCA models have been developed. In this section, we review these state-of-art popular DCA models.

A. Multidimensional Independent Component Analysis

Hyvarinen and Hoyer presented a special case of MICA which they called independent subspace analysis(ISA)[4]; there the dependence within a k-tuple is explicitly modelled enabling the authors to propose better algorithms without having to resort to the problematic multidimensional density estimation. A different extension of ICA is given by topographic ICA [5], where dependencies between all components are assumed.

The goal of a general ISA is the decomposition of an arbitrary random vector $\mathbf{x}$ into independent components. If $\mathbf{x}$ is to be decomposed into one-dimensional components, this coincides with ordinary ICA. Similarly, if the independent components are required to be of the same dimension k, then this is denoted of fixed group size k or simply k-ISA. So 1-ISA is equivalent to ICA.

Let $k,n\in N$ such that k divides n. We call an n-dimensional random vector $\mathbf{y}$ k-independent if the k-dimensional random vectors

$$(y_1,\cdots,y_k)^T,\cdots,(y_{n-k+1},\cdots,y_n)^T$$

are mutually independent. A matrix $\mathbf{W}\in Gl(n,\mathbb{R})$ is called a k-ISA of an n-dimensional random vector $\mathbf{x}$ if $\mathbf{Wx}$ is k-independent. If k = 1, this is the same as ordinary ICA.

Obvious indeterminacies are, similar to ordinary ICA, invertible transforms in $Gl(n,\mathbb{R})$ in each tuple as well as the fact that the order of the independent k-tuples is not fixed. So, define for $r,s=1,\cdots,n/k$ the (r,s) sub-k-matrix of $\mathbf{W}=(w_{ij})$ to be the $k\times k$ submatrix

$$(w_{ij})_{\substack{i=rk,\cdots,rk+k-1\\ j=sk,\cdots,sk+k-1}}$$

that is the $k\times k$ submatrix of $\mathbf{W}$ starting at position (rk,sr). A matrix $\mathbf{L}\in Gl(n,\mathbb{R})$ is said to be a k-scaling and permutation matrix if for each $r=1,\cdots,n/k$ there exists precisely one s with the (r,s) sub-k-matrix of $\mathbf{L}$ to be nonzero, and such that this submatrix is in $Gl(n,\mathbb{R})$, and if for each $s=1,\cdots,n/k$ there exists only one r with the (r,s) sub-k-matrix satisfying the same condition. Hence, if $\mathbf{y}$ is k-independent, also $\mathbf{Ly}$ is k-independent.

Two matrices $\mathbf{A}$ and $\mathbf{B}$ are said to be k-equivalent, $\mathbf{A}\sim_k\mathbf{B}$, if there exists such a k-scaling and permutation matrix $\mathbf{L}$ with $\mathbf{A}=\mathbf{BL}$. As stated above, given two matrices $\mathbf{W}$ and $\mathbf{V}$ with $\mathbf{W}^{-1}\sim_k\mathbf{V}^{-1}$ such that one of them is a k-M ICA of a given random vector, then so is the other. We will show that there are no more indeterminacies of MICA.

As usual MICA can solve the MBSS problem

$$\mathbf{x}=\mathbf{As}$$

where $\mathbf{A}\in Gl(n,\mathbb{R})$ and $\mathbf{s}$ is a k-independent n-dimensional random vector. Finding the indeterminacies of MICA then shows that $\mathbf{A}$ can be found except for k-equivalence (separability), because if $\mathbf{x}=\mathbf{As}$ and $\mathbf{W}$ is a demixing matrix such that $\mathbf{Wx}$ is k-independent, then $\mathbf{WA}\sim_k\mathbf{I}$, so $\mathbf{W}^{-1}\sim_k\mathbf{A}$ as desired.

However, for the proof we need one more condition for $\mathbf{A}$: We call $\mathbf{A}$ k-admissible if for each $r,s=1,\cdots,n/k$ the (r,s) sub-k-matrix of $\mathbf{A}$ is either invertible or zero. Note that this is not a strong restriction—if we randomly choose $\mathbf{A}$ with coefficients out of a continuous distribution, then with probability one we get a k-admissible matrix, because the non-k-admissible matrices belong to $\mathbb{R}^{n^2}$ lie in a sub-manifold of dimension smaller than n^2.

Theorem 2. *(Separability of MICA)*[21] Let $\mathbf{A}\in Gl(n,\mathbb{R})$ and $\mathbf{s}$ a k-independent n-dimensional

random vector having no Gaussian k-tuple $(s_{rk},\cdots,s_{rk+k-1})^T$. Assume that $\mathbf{A}$ is k-admissible.

If $\mathbf{As}$ is again k-independent, then $\mathbf{A}$ is k-equivalent to the identity. For the case $k=1$ this is linear BSS separability because every matrix is 1-admissible.

The MICA method used for biomedical signals are shown in Figure 1.

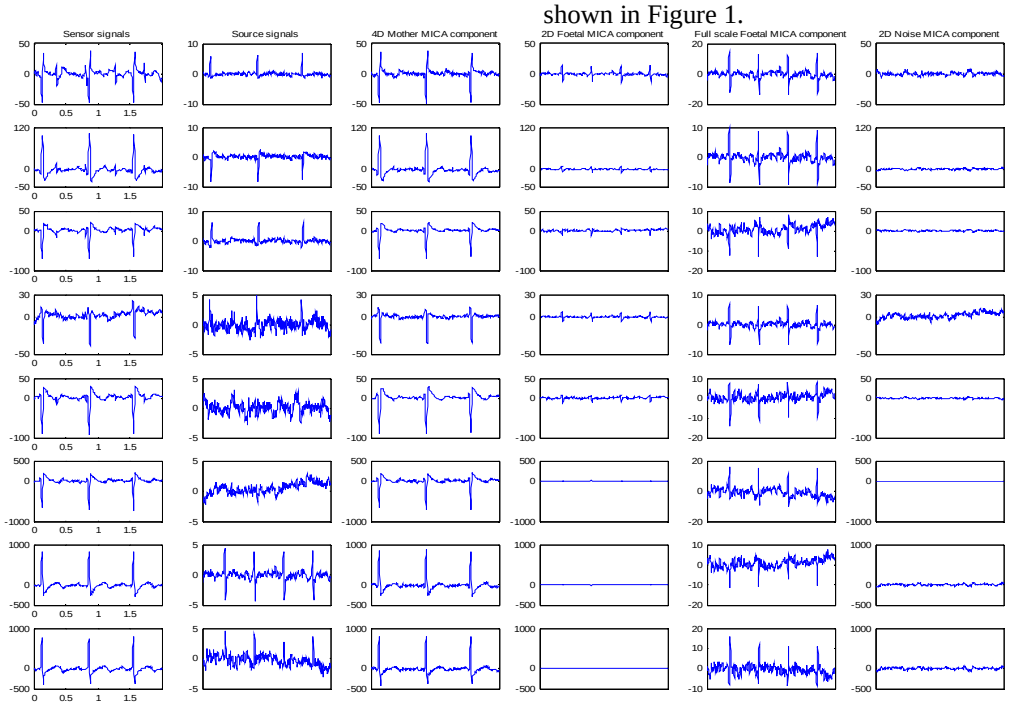

Figure 1. The results using MICA method to DAISY data on http://www.tsi.enst.fr/~cardoso/RRicassp98.htm .

B. Variance dependent BSS model

Among many extensions of the basic ICA models, several researchers have studied the case where the source signals are not independent. The dependencies either need to be exactly known beforehand, or they are simultaneously estimated by the algorithms. Recently, a novel idea called double-blind approach was introduced by Hyvarinen et.al.[6,7]. In contrast to previous work, their method requires no assumption on the distributions of the sources and no parametric model of the dependencies between the components. They simply assume that the sources are dependent only through their variances and that the sources have temporal correlation. In the Topographic ICA[8], the dependencies of the sources are also caused only by their variances, but in contrast to the double blind case, they are determined by a prefixed neighborhood relation. It should be noted that for such dependent component models identifiability results have not been theoretically established so far, while identifiability of multidimensional ICA was proven by Theis [21].

A statistical basis of ICA was established by Amari and Cardoso[22] which pointed out that the ICA model is an example of semiparametric statistical models and studied estimating functions for it. In particular, they showed that the quasi maximum likelihood(QML) estimation and the natural gradient learning give a correct solution regardless of the true source densities which satisfy certain mild conditions. [7] extend their approach to the BSS problem. By investigating estimating functions for the model, they show that many of ICA algorithms based on the independence assumption can achieve consistent solutions in a local sense, even if there exist variance dependencies, which is astonishing and seems somewhat counterintuitive. For a few algorithms, even global consistency has been proven by different principles (for example, [11]). Nevertheless, our result goes beyond existing ones, because it covers most types of BSS algorithms and can give asymptotic distributions. The main message of this paper is that most ICA algorithms can be proven to be consistent in our framework although the data is not independent. So they must effectively use some concept beyond independence. Thus our consistency results indicate that separation can be done based only on normalized sources which are adjusted to have stationary variances and is not affected by the dependent activity levels.

Hyvarinen formalized the probabilistic framework of variance-dependent blind separation [7]. [6] assumes that each source signal $s_i(t)$ is a product of non-negative activity level $v_i(t)$ and underlying i.i.d. signal $z_i(t)$, that is, $s_i(t)=v_i(t)z_i(t)$. We remark that the sequences of the vectors $\mathbf{s}=(s_1,\cdots,s_n)^T$, $\mathbf{v}=(v_1,\cdots,v_n)^T$ and $\mathbf{z}=(z_1,\cdots,z_n)^T$ are considered as multivariate random processes in this subsection. In practice, the activity levels $v_i(t)$ are often dependent among different signals and each observed signal is expressed as

$$x_i(t) = \sum_{j=1}^{n} a_{ij} v_j(t) z_j(t), \qquad i = 1, \cdots, n \tag{3}$$

where $v_i(t)$ and $z_i(t)$ satisfy: 1) $v_i(t)$ and $z_j(t')$ are independent for all i, j, t, t'; 2) each $z_i(t)$ is i.i.d. in time for all i, the random vector $\mathbf{z} = (z_1, \cdots, z_n)^T$ is mutually independent; 3) $z_i(t)$ have zero mean and unit variance for all i.

No assumption on the distribution of z_i is made except 3). Regarding the general activity levels v_i's, $v_i(t)$ and $v_j(t)$ are allowed to be statistically dependent, and furthermore, no particular assumption on these dependencies is made (double blind situation). We refer to this framework as the variance dependent BSS model in this paper. As stated in the assumption 2) above, the normalized signals z_1 and z_2 are mutually independent.

However, since the sequences z_1 and z_2 are multiplied by extremely dependent activity levels v_1 and v_2, respectively, the short-term variance of the source signals s_1 and s_2 are highly correlated.

Theorem 3. Assume that the signals $x_i(t)$ are generated as described in Eq. (3), and that the signals are preprocessed by spatial whitening to give the multidimensional signal $\mathbf{z}(t)$. Define the objective function:

$$J(\mathbf{W}) = \sum_{i,j} [\mathrm{cov}(\mathbf{w}_i^T \mathbf{z}(t))]^2, [\mathbf{w}_j^T \mathbf{z}(t - \Delta t)]^2)]^2, \tag{4}$$

where $\mathbf{W} = (\mathbf{w}_1, \cdots, \mathbf{w}_n)^T$ is constrained to be orthogonal, and the lag Δt is nonzero. Assume that the matrix $\mathbf{K}$ defined as

$$\mathbf{K}_{ij} = \mathrm{cov}(s_i^2(t), s_j^2(t - \Delta t)) \tag{5}$$

has full rank. Then, the objective function J is (globally) maximized when $\mathbf{WA}$ equals a signed permutation matrix, i.e. the $\mathbf{w}_i^T \mathbf{z}(t)$ equal the original sources $s_i(t)$ up to random signs. Convergence of the variance dependent BSS algorithm for artificially generated data are shown in Figure 2.

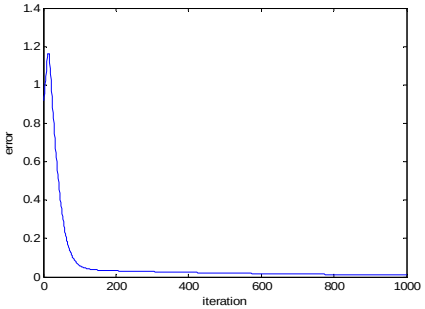

Figure 2. Convergence of the variance dependent BSS algorithm for artificially generated data. Vertical axis: error, horizontal axis: iteration count.

C. Subband Decomposition ICA(SDICA)

Subband decomposition ICA(SDICA), an extension of ICA, assumes that each source is represented as the sum of some independent subcomponents and dependent subcomponents, which have different frequency bands. SDICA model [10-13] can considerably relax the assumption regarding mutual independence between the original sources by assuming that the wide-band source signals are generally dependent but some narrow-band subcomponents of the sources are independent.

In general, in order to solve the SDICA problem, The first thing is to apply a filter to the observations to allow the frequency bands of the independent subcomponents to pass through and then apply the standard ICA algorithms to these filtered signals. In [12], the subband is selected by a priori knowledge or by comparing some simple statistical measures (such as l_p-norm, p =1, or 0.5, or kurtosis) of the subband signals. Generally these selection methods seem to be arbitrary if we do not have the prior information on the subband of the independent subcomponents. In [10], an additional assumption that at least two groups of subcomponents are statistically independent is incorporated, so that the true mixing matrix can be recovered. This assumption may help solve some practical BSS problems. However, it does not necessarily hold. And even with this assumption, it remains a problem to divide the optimal subbands. Moreover, these methods cannot recover the source-independent subcomponents. In some cases, the source-independent subcomponents are of interest, and the dependent subcomponents should be removed. For example, the dependent subcomponents may denote the power supply, which yields a sinusoidal inference.

Zhang et.al. discuss the validity of the existing methods for SDICA and the conditions for separability of the SDICA model [11]. They propose an adaptive method for SDICA, called band-selective ICA (BS-ICA). This method can automatically select the frequency band in which the subcomponents of the original sources are most independent, and consequently the mixing matrix and a filtered version of the source-independent subcomponents can be estimated. In order to do that, a linear filter on each observation is applied, followed by a linear demixing stage. The parameters in the filter and the demixing matrix are adjusted by minimizing the mutual information between the outputs. By incorporating some penalty term, the prior knowledge on the independent subcomponents can be taken into account.

The overcomplete model is also considered in [11]. In overcomplete DCA model, the number of the original independent sources is greater than that of the observations, so standard ICA algorithms cannot cope with this problem. [11] is concerned with the overcomplete ICA problems in which there exists a subset of sources such that each source in this subset has some frequency band outside the frequency bands of the sources not in this subset. The relationship between such overcomplete ICA problems and SDICA is addressed. Based on the relationship, BS-ICA can also be exploited to solve this kind of overcomplete ICA problem.

The key idea in SDICA is the assumption that the wide-band source signals can be dependent; however, only some of their narrow-band subcomponents are independent [12]. In other words, all sources s_i are not necessarily independent, but can be represented as the sum of several subcomponents as

$$s_i(t) = s_{i,1}(t) + s_{i,2}(t) + \cdots + s_{i,L}(t) \quad (6)$$

where $s_{i,k}(t)$, $k = 1, 2, \cdots, L$ are narrow-band subcomponents. And the subcomponents are mutually independent for only a certain set of *k*; more precisely, we assume that the subcomponents with *k* in this set are spatially independent stochastic sequences. The observations are still generated from the sources s_i according to equation (2). Here we assume that the number of sources is equal to that of the observations and that the observations are zero mean. Similar to ICA, the goal of SDICA is to estimate the mixing matrix, the original sources, and the source-independent subcomponents if possible.

SDICA can be performed by the structure in Figure 3. As the first step, we apply a filter $h(t)$ to filter out the dependent subcomponents of the sources. Suppose $h(t)$ exactly allows one independent subcomponent, say, the *k*th subcomponent $s_{i,k}(t)$ pass through. Let $\mathbf{s}^{(k)}(t) = [s_{1,k}(t), \cdots, \ . \ s_{n,k}(t)]^T$. Then the filtered observations

$$h(t) * \mathbf{x}(t) = \mathbf{A}[h(t) * (\mathbf{s})(t)] = \mathbf{A}\mathbf{s}^{(k)}(t).$$

Therefore, in the second step, we just need to apply an ICA algorithm to the filtered observations, and we can obtain the demixing matrix $\mathbf{W}$ associated with the mixing matrix $\mathbf{A}$.

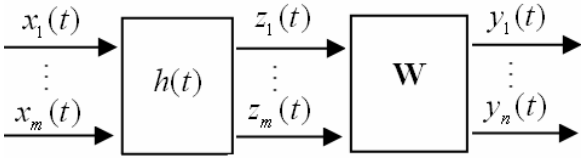

Figure 3. The structure to perform SDICA.

In the existing methods for SDICA, the frequency subband in which source sub-components are independent is determined by either some a priori information or exploiting the stronger assumption that at least two of the subcomponents are statistically independent. In practice, ICA is mainly used for BSS, so the exact a priori information on $h(t)$ is usually unavailable. The assumption that at least two of the subcomponents are statistically independent is not necessarily true, and the design of the optimal filter $h(t)$ for good performance remains a problem. It is therefore very useful to develop a method that adaptively estimates the optimal filter $h(t)$, such that the source-independent subcomponents pass through it and the dependent subcomponents are attenuated, and consequently both the mixing matrix A and the source-independent subcomponents can be recovered. Band-selective ICA (BS-ICA) is such a method [11].

Separability of SDICA

In last section we briefly reviewed the existing methods for SDICA. Now we provide the theoretical foundations that sustain these methods. Furthermore, we discuss why it is possible to adaptively estimate the SDICA separation system $\{h(t), \mathbf{W}\}$ without a priori knowledge on $h(t)$.

Proposition 1. Let $\mathbf{x}(t)$ be the observations in the SDICA model. Under assumptions 1 and 2, the outputs of the SDICA separation system in Figure 3, $y_i(t)$, are spatially independent stochastic sequences if and only if the filter $h(t)$ filters out the dependent subcomponents $s_{i,D}(t)$ and $\mathbf{WA}$ is a generalized permutation matrix.

Proposition 2. Let $\mathbf{x}(t)$ be the observations in the SDICA model. Under assumptions 1 to 5, the outputs of the SDICA separation system, $y_i(t)$, are instantaneously independent if and only if the filter $h(t)$ filters out the dependent subcomponents $s_{i,D}(t)$ and $\mathbf{WA}$ is a generalized permutation matrix.

Assumptions 1 to 5 [11] are generally not very restrictive. And it is important to emphasize that it is possible for proposition 2 to be true even when some of the assumptions are violated. According to proposition 2, the SDICA model generated observations can be separated by the SDICA separation system in Figure 3 and the filter $h(t)$ and the demixing matrix $\mathbf{W}$ in the SDICA system can be obtained by making $y_i(t)$ mutually independent using the Band-selective method.

In order to investigate the performance of the SDICA BSS algorithm, the simulation is shown below, the source signals are selected as harmonic signals as proposed in [30], the four source signals are generated as follows:

$$s_1(t) = 0.5\cos(2\pi t N_1/1024 + 1.9);$$
$$s_2(t) = 0.9\cos(2\pi t N_2/1024 + 0.8);$$
$$s_3(t) = 1.6\cos(2\pi t N_3/1024 + 0.1);$$
$$s_4(t) = 2.0\cos(2\pi t N_4/1024 + 0.2).$$

where $N_1 = 50$, $N_2 = 170$, $N_3 = 290$, $N_4 = 410$.

The source signals were contaminated with additive moving average(MA)(2) noise, and the signal-to-noise ratio(SNR) is 0db. The noise model is

$$n(t) = e(t) - 0.45e(t-1) + 0.95e(t-2)$$

where $e(t) \sim N(0,1)$.

The mixed harmonic signals in noise is generated by the ICA based harmonic retrieval models [30] and $l_1 = 2$, $l_1 = 4$, $l_1 = 6$, $l_1 = 8$. The source harmonic signals, mixtures with no noise, and their Fourier transformation are displayed in figure3(a) and figure3 (b), respectively.

The separation results of adaptive filter based method is given in figure3(c). Because of the algorithm by the adaptive filter based method is affected by the noise

greatly, we also give the separated results in the case of no noise which is given in figure3(c).

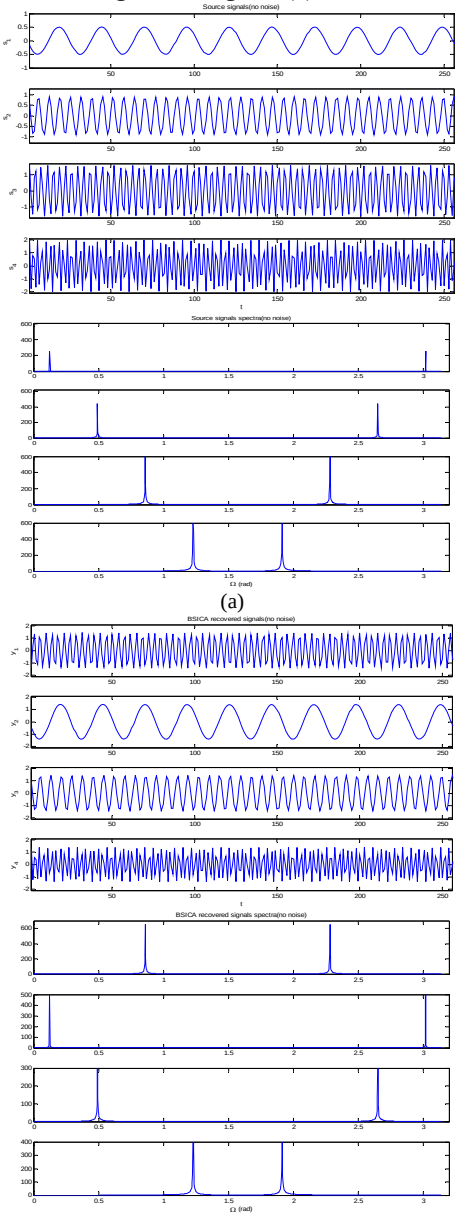

Figure 4. (a)The source signals and their FFT; (b)The simulation results using adaptive filter based method with no noise. The data length is 1024, we show only the first 256 for clear comparison with source signals.

D. Maximum non-Gaussianity Method

In ICA applications, non-Gaussianity measures are used based on the following fundamental idea: the outputs of a linear mixing process that preserves variances, have higher entropies than the inputs. This general statement can be precisely expressed in mathematical terms when source signals are mutually independent as a consequence of central limit theorem (CLT) which tell us that the linear mixture of *n* independent signals with finite variances will became asymptotically Gaussian (as *N* grows towards ∞). Moreover, Donoho, in his classical paper on blind deconvolution [25] has shown that any finite linear combinations of independent random variables are "more nearly" Gaussian than the individual components.

Of course, when signals (variables) are dependent, the classical CLT does not hold and we cannot be sure that maximum NG(MaxNG) method will reach to the sources. Moreover, in [14], a particular case where this method fails is presented, but fortunately this is not the case in most of real world scenarios. They present many experimental results showing that the method is useful in many cases, even when the independence of sources is relaxed and also we show that MaxNG performs always better than MinMI. Besides, there are many experimental evidences from other authors, in the framework of blind deconvolution, where independence of input variables is not required [26]. Particularly, in [25] there is an example of dependent variables that can be reconstructed using the minimum entropy method.

The theoretical conditions to be satisfied by sources in order to assure MaxNG to work, may be based on those that allow us to generalize the CLT to special dependent variables. Enormous amount of work have been published since long time ago establishing different sufficient conditions for this generalization. Anyway the characterization of the type of dependence of variables that guarantees the generalization of CLT is cumbersome and remains as an open issue at present.

The definition of the MaxNG method is as follows:

Definition 4.[14]. The maximum NG (minimum entropy) method consists of searching for the linear combinations of mixtures that give source estimates with maximum non-Gaussian (minimum entropy) distributions restricting the space of search to the unit-variance signals space. More specifically, sources are estimated through relation $\mathbf{y} = \mathbf{D}\mathbf{x}$ over the space of invertible separating matrices $\mathbf{D}$ providing signals $y_0, \cdots, y_{M-1}$ with unit-variances (which is equivalent to imposing the covariance matrix $\mathbf{R}_{\mathbf{yy}} = \mathbf{D}\mathbf{R}_{\mathbf{xx}}\mathbf{D}^T$ to have ones in its main diagonal).

Let us now introduce a natural measure of NG based on the L^2-Euclidean distance of an estimated pdf to the normal (Gaussian) pdf. Considering a continuous random variable y with zero-mean and unit-variance, we define our NG measure of a pdf p_y denoted by $\Gamma(p_y)$, as following:

$$\Gamma(p_y) = \int [\Phi(y) - p_y(y)]^2 dy \,, \tag{8}$$

where the integral is defined in Lebesgue sense and is taken on all the range of variable y, and $\Phi(y)$ is the Gaussian pdf:

$$\Phi(y) = N(0,1) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}y^2\right) \tag{9}$$

Clearly, Eq. (8) is the square of the distance between functions $\Phi(y)$ and $p_y(y)$ in L^2 . We restrict our analysis to the case where $p_y(y)$ is a continuous function for simple mathematical treatment. Note that zero is a

lower bound of $\Gamma(p_y)$ and is attained when $p_y(y)=\Phi(y)$.

Suppose that our data are samples of the random variable y: $y(0), y(1), \cdots, y(N)$. If the number of samples is large enough, generally, we can obtain a good estimation of the unknown pdf $\hat{p}_y(y)$ using the non-parametric technique, namely Parzen window, with a Gaussian kernel [27] as follows:

$$\hat{p}_y(y)=\frac{1}{Nh}\sum_{i=0}^{N-1}\Phi\left(\frac{y-y(i)}{h}\right), \quad (10)$$

where *N* is the number of samples, *h* is a parameter which affects the width and height of the windows functions in the summation, and $\Phi(y)$ is the Gaussian window or kernel as was defined in (9). In the following it will show that the selection of Gaussian kernel allows us to derive some useful relationships.

Working with Eq. (8) we divide our NG measure in three parts as follows:

$$\Gamma(p_y)=\underbrace{\int\Phi(y)^2dy}_{\frac{1}{2\sqrt{\pi}}}\underbrace{-2\int\Phi(y)p_y(y)dy}_{\Gamma_1(p_y)}+\underbrace{\int p_y^2(y)dy}_{\Gamma_2(p_y)} \quad (11)$$

The first term of the right hand side in Eq. (11) can be analytically calculated and it takes $1/2\sqrt{\pi}$ value. Only the second $(\Gamma_1(p_y))$ and third $(\Gamma_2(p_y))$ terms are dependent on the pdf p_y.

Replacing $p_y(y)$ by $\hat{p}_y(y)$ given by Eq. (10) and using properties of the convolution of Gaussian functions, we arrive finally to the following formulas [14]:

$$\Gamma_1(\hat{p}_y)=\frac{-2}{N\sqrt{h^2+1}}\sum_{i=0}^{N-1}\Phi\left(\frac{y(i)}{\sqrt{h^2+1}}\right), \quad (12)$$

$$\Gamma_2(\hat{p}_y)=\frac{1}{N^2h\sqrt{2}}\sum_{i=0}^{N-1}\sum_{j=0}^{N-1}\Phi\left(\frac{y(j)-y(i)}{\sqrt{2}h}\right). \quad (13)$$

We propose to set $h=1.06\times N^{-1/5}$ which is the value that guarantees to attain the minimum mean integrated square error (MISE) in the pdf estimation for a Gaussian kernel. This value is optimum for the pdf estimation and it is reasonable to expect good estimates of our non-Gaussian distance as well.

While the Euclidean distance is a very common way to measure distances between functions and it was already used in some separation algorithms (see [28] where a calculation of this distance is based on orthogonal polynomial expansions), the set of formulas (12) and (13), which are obtained using Parzen windows technique are new.

There is also an interesting direct connection between term $\Gamma_2(p_y)$ in our Eq. (13) and Renyi entropy of order 2 $H_{R_2}(y)$, which is defined as $H_{R_2}(y)=-\log\int p_y^2(y)dy$ and was used before in BSS and feature extraction (see [29]).

E. Spectral(Wold Decomposition) Method[25]

For signal model:

$$\mathbf{x}(t)=\mathbf{A}\mathbf{s}(t)+\mathbf{n}(t),$$

there are some assumptions:

1) Each element of $\mathbf{s}(t)$ is a zero-mean stationary process.

2) $\mathbf{n}(t)$ is zero-mean stationary, white random vector process, independent of the source signals.

3) A has full column rank, i.e. rank(**A**)=*n*, otherwise **A** can have any unknown form.

It must be emphasized that we do not impose any assumption about the independence or uncorrelation of source signals. In other words, source signals can be correlated, and only the following assumption is considered:

4) Source signals are jointly stationary.

The aim in BSS is to recover the source signals from the observations, and for attaining this target we try to identify the mixing matrix **A**. Two indeterminacies are exist as the classic BSS problem.

F. Time-Frequency Method[18,19]

Assumption 1.

1) The mixing matrix **A** is such that $a_{ij}\neq 0, \forall i, j$.

2) The power of each source is non-negligible at least at some times t.

Assumption 2. For each source s_i there exist some adjacent TF windows (t_j,ω_k) centered on time t_j and angular frequency ω_k where only s_i occurs, i.e. where: $S_l(t_j,\omega_k)\ll S_i(t_j,\omega_k), \forall l\neq i$

Assumption 3. When several sources occur in a given set of adjacent TF windows they should vary so that $\alpha(t,\omega)$ does not take the same value in all these windows. Especially,

1) At least one of the sources must take significantly different TF values in these windows;

2) The sources should not vary proportionally and

$$\alpha(t,\omega)=\frac{\sum_{m=1}^{N}a_{1m}S_m(t,\omega)}{\sum_{m=1}^{N}a_{2m}S_m(t,\omega)}.$$

Assumption 4. There exists a TF area (Γ_q,ω_k) where only source s_i occurs and, $\forall(t_j,\omega_k)\in(\Gamma_q,\omega_k)$, $|N_1(t_j,\omega_k)|\ll|S_i(t_j,\omega_k)|$ and $|N_2(t_j,\omega_k)|\ll|S_i(t_j,\omega_k)|$.

IV. CONCLUSIONS

In most BSS methods, the source signals are assumed to be statistically independent. BSS based on such a model is called ICA. However, the independence property of sources may not hold in some real-world situations, especially in biomedical signal processing and image processing, and therefore the standard ICA cannot give the expected results. Some extended data models have been developed to relax the independence

assumption in the standard ICA model, we call them DCA as a whole. DCA as an extended ICA model has more application than ICA and received more and more attentions during the last several years in the study of signal processing, neural network and applications. As two important BSS methods, ICA and DCA have many relations from the models to the optical algorithms. In this paper, we describe the overall basic DCA model in detail, moreover, the relationships between ICA and these DCA models have also been shown. The separateness and uniqueness is discussed of some special DCA models too. At last, the state-of-art DCA algorithms are overviewed from different theory foundations, such as, multidimensional ICA, variance dependent BSS, subband decomposition ICA, maximum non-Gaussianity method, Wold decomposition method and time-frequency method are constructed for the BSS problem in theories and some simulations of these algorithms are also exhibited for different applications.

ACKNOWLEDGMENT

This work is partially supported by National Natural Science Foundation of China(Grant No.60672049) and the Science Foundation of Henan University of Technology(Grant No.08XJC027).

REFERENCES

[1] A. Cichocki, S. Amari, *Adaptive Blind Signal and Image Processing: Learning Algorithms and Applications*, John Wiley&Sons, New York, 2002.
[2] A. Hyvarinen, J. Karhunen, E. Oja, *Independent component analysis*, John Wiley &Sons, New York, 2001.
[3] J.F. Cardoso, "Multidimensional independent component analysis", *In Proc. of the IEEE 1998 Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP'98)*, pp:1941-1944, Seattle, WA, USA, IEEE, 1998.
[4] A. Hyvarinen, P. Hoyer, "Emergence of phase and shift invariant features by decomposition of natural images into independent feature subspaces", *Neural Computation*, vol.12, no.7, pp.1705-1720, 2000.
[5] H. Valpola, M. Harva, J. Karhunen, "Hierarchical models of variance sources", Signal Processing, vol.84, no.2, 2003.
[6] M. Kawanabe, K.R. Muller, "Estimating functions for blind separation when sources have variance dependencies", *Journal of Machine Learning Research*, vol.6, pp. 453-482 , 2005.
[7] A. Hyvarinen, J. Hurri, "Blind separation of sources that have spatiotemporal variance dependencies". *Signal Processing*, 84, 2004. 247–254.
[8] A.Hyvarinen, P.O.Hoyer, M. Inki, "Topographic Independent Component Analysis", *Neural Computation*, vol.13, no.7, pp.1527-1558, 2001.
[9] F.R. Bach, M.I. Jordan, "Kernel Independent Component Analysis", *Journal of Machine Learning Research*, vol.3, pp.1-48, 2002.
[10]T.Tanaka, A. Cichocki. "Subband decomposition independent component analysis and new performance criteria". *In Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP'04)*, volume 5, pages 541–544, 2004.
[11]K. Zhang, L.W. Chan, "An Adaptive Method for Subband Decomposition ICA", *Neural Computation*, vol.18, no.1, pp.191-223, 2006.
[12]A. Cichocki, P. Georgiev. "Blind source separation algorithms with matrix constraints". *IEICE Transactions on Information and Systems, Special Session on Independent Component Analysis and Blind Source Separation*, E86-A (1), 522–531, 2003.
[13]A. Cichocki, T.M. Rutkowski, K. Siwek, "Blind signal extraction of signals with specified frequency band". *In Neural Networks for Signal Processing XII: Proceedings of the 2002 IEEE Signal Processing Society Workshop*. Piscataway, NJ: IEEE. pp. 515–524, 2002.
[14]C.F. Caiafa, A.N. Proto, "Separation of statistically dependent sources using an L^2-distance non-Gaussianity measure", *Signal Processing*, vol.86, no.11, pp.3404-3420, 2006.
[15]F.S. Wang, H.W. Li, R. Li, "Novel NonGaussianity Measure Based BSS Algorithm for Dependent Signals", *Lecture Notes in Computer Science*, vol.4505, pp.837-844, 2007.
[16]C.F. Caiafa, E. Salerno, A.N. Protoa, L. Fiumi, "Blind spectral unmixing by local maximization of non-Gaussianity", *Signal Processing*, vol.88, no.1, pp.50-68, 2008.
[17]M.R. Aghabozorgi, A.M. Doost-Hoseini, "Blind separation of jointly stationary correlated sources", *Signal Processing*, vol.84, no.2, pp.317-325, 2004.
[18]F. Abrard, Y. Deville, "A time-frequency blind signal separation method applicable to underdetermined mixtures of dependent sources", *Signal Processing*, vol.85, no.7, pp.1389-1403, 2005.
[19]R. Li, F.S. Wang, "Efficient Wavelet Based Blind Source Separation Algorithm for Dependent Sources", *Advances in Soft Computing*, Springer Press, vol.40, pp.431-441, 2007.
[20]P.Comon, "Independent component analysis, a new concept?" *Signal Processing*, vol.36, no.3, pp.287-314, 1994.
[21]F.J. Theis, "Uniqueness of complex and multidimensional independent component analysis", *Signal Processing*, vol.84, no.5, pp. 951-956, 2004.
[22]S. Amari, J.F. Cardoso. Blind source separation-semiparametric statistical approach. IEEE Trans. on Signal Processing, 45(11):2692–2700, 1997.
[23]A. Hyvarinen. "Independent component analysis for time-dependent stochastic processes". *In Proc. Int. Conf. on Artificial Neural Networks (ICANN'98)* Skovde, Sweden, pp. 541–546, 1998.
[24]O. Yilmaz, S. Rickard, "Blind separation of speech mixtures via timefrequency masking". *IEEE Transactions on Signal Processing*, vol. 52, pp.1830–1847, 2004.
[25]D. Donoho, "On minimum entropy deconvolution", *Applied Time Series Analysis II*, Academic Press, New York, pp.565–609,1981.
[26]R.A. Wiggins, "Minimum entropy deconvolution", *Geoexploration*, vol. 16, pp.21–35, pp.1978.
[27]R.O. Duda, P.E. Hart, D.G. Stork, *Pattern Classification*, second ed., Wiley, New York, 2000.
[28]T. Yokoo, B.W. Knighty, L. Sirovich, "L^2 de-Gaussianisation and independent component analysis", *The Fourth International Symposium on Independent Component Analysis and Blind Signal Separation (ICA 2003)*, April 2003, Nara, Japan.
[29]K. Torkkola, "Feature extraction by non-parametric mutual information maximization", *Journal of Machine Learning Research*, vol.3, pp.1415–1438, 2003.

Rui LI was born in Henan, China, in 1979. She received her M.S. degree in Computation Mathematics in 2008 from Zhengzhou University. Currently, she is a lecture in Henan University of Technology. Her research interests include blind signal processing, intelligence optimization.

Hongwei LI was born in Hunan, China, in 1965. He received his Dr. degree in Applied Mathematics in 1996 from Beijing University. Currently, he is a professor in China University of Geosciences, China. His research interests include statistical signal processing, nonstationary signal processing, blind signal processing.

Fasong WANG was born in Henan, China, in 1979. He received his M.S. degree in Applied Mathematics in 2005 from China University of Geosciences, China. He is currently working as an engineer at China Electronics Technology Group Corporation, China. His current research interests include blind signal processing, time-frequency analysis and their applications. He has published more than thirty journal and conference papers on these topics.